PROBLEMS OF ANALYTICAL STATISTICS

Academician Yu. V. LINNIK

STATISTICAL PUBLISHING SOCIETY
CALCUTTA 700 035

© Statistical Publishing Society
204/1 B. T. Road, Calcutta 700 035

First Published : *August 1975*

Printed at Eka Press, 204/1, Barrackpore Trunk Road, Calcutta 700 035.

Math
QA
276
.L56

PREFACE

This publication is an enlarged version of the lectures given at the Statistical Institute of the Paris University at the end of 1966. The aim of the lectures was to demonstrate the useful applications of the theory of functions of one and many complex variables to the statistical theory of testing and estimation which were found recently by many authors.

The book is a continuation of the author's monograph "Statistical problems with nuisance parameters", Moscow, 1966 (English translation was published by the *Amer. Math. Soc.* in 1968). For many lemmas and particularly for purely analytic calculations the reader is referred to this monograph. The attention of the lectures was mainly concentrated on the statistical spirit of the problems; the analytic lemmas which cannot be found in the monograph mentioned above, are proved here in details. The elementary knowledge of modern analysis including the elementary properties of the L. Schwartz distributions is supposed. As to the knowledge of statistics it is supposed that the reader is familiar with material as, for instance, in a book by D. Dugué "Traité de statistique pure et appliquée", I, II, Paris, Masson et Cie, 1958.

It is a pleasure for the author to express his gratitude to Professors D. Dugué and S. Mandelbroit for their interest and valuable advice, to Mr. B. Morucci and Mr. Dufieux for their friendly remarks, critique and corrections of the text and to Miss C. Mouzon and S. Philippe for their help in preparing the manuscript.

Yu. V. LINNIK

The University of Iowa
Libraries

CONTENTS

	PAGE
PREFACE	v
INTRODUCTION	ix
Exponential families. Sufficient and partilaly sufficient statistics	1
Applications of sufficient statistics. Independent statistics. ..	7
On "unlinking" independent statistics	16
Problems with nuisance parameters. Verifiable functions. Similar tests	21
Cotests : ideals of cotests. The V. Palamodov theorem ..	29
Theorem of V. Palamodov (*continuation*)	35
Application of the V. Palamodov theorem to the Behrens-Fisher problem	41
Properties of homogeneous similar tests for the Behrens-Fisher problem	45
On the Fisher-Welch-Wald test	53
Construction of a non-randomized homogeneous similar test for the Behrens-Fisher problem	61
Verfiable functions	67
Verifiable polynomials. Sequentially verifiable functions ..	77
On unbiased estimation for incomplete exponential families with polynomial relations	85
Some properties of estimation of a location parameter. ..	95
Minimax detection of a discrete signal disturbed by a normal noise	103

INTRODUCTION

The analytical statistics can be considered as applications of modern analysis to statistical problems such as hypotheses testing (in both fixed-size and sequential sampling) estimation and other branches of statistical decision theory. It should be noted that the general statistical decision theory, though well developed in its general outline, presents many difficulties when one tries to elaborate on its more fine and interesting structures.

Thus, for example, the problem of optimal sequential choice among finite number of statistical hypotheses has not yet been solved except in the simplest cases. Only in rather simple situations applications of modern analysis, in the first place of the theory of analytic functions of many variables, can be carried out successfully and we shall try to demonstrate here some of these applications. The special problems which will be considered are theory of similar tests, in particular the Behrens-Fisher problem, estimation theory for the exponential families, theory of minimax tests in multivariate analysis.

The application of the theory of functions of one and many complex variables to probability theory began in papers by H. Cramér, D. Dugué, D. Raikov and others; as to statistics we present here some new results obtained in particular by the Soviet scholars A. Kagan and V. Palamodov.

We suppose that the reader is acquainted with the basic concepts of statistics, in particular with the theory of sufficient statistics. For definitions and general theorems we refer the reader to the well known monograph "Testing Statistical Hypotheses" by E. L. Lehmann.

CHAPTER ONE

EXPONENTIAL FAMILIES. SUFFICIENT AND PARTIALLY SUFFICIENT STATISTICS

The concept of the exponential families of distribution functions can be traced to H. Poincaré and L. Boltzmann. These families contain many distributions of unquestionable theoretical and practical importance, for example, the normal distributions of any dimension, the distributions of Pearson type III etc. The most natural view point on these families was formulated by the soviet mathematician N. Chentzov [2].

N. Chentzov was the first who introduced into statistics the concept of categories by a proper definition of the objects and morphisms of statistics. It can be noticed that the N. Chentzov morphisms are closely connected with the Fredholm equations of type I which we shall discuss in connection with the problems of minimax tests in multivariate analysis. A special metric (invariant metric with respect to the Chentzov morphisms) is introduced and then the exponential families are represented as the geodesic lines with respect to this metric. Unfortunately we cannot discuss here the details of the interesting theory of N. Chentzov and so refer the reader to his note [2].

Let us consider now the simplest case of the exponential families.

Let $X = E_k$ be the Euclidean space with the σ-algebra $\mathcal{A}$ of Borel sets and let π be a σ-finite measure defined on $\mathcal{A}$. Consider the family of densities $\{p_\theta(x)\}$ with respect to π of the following form :

$$p_\theta(x) = C(\theta)h(x) \exp\left(\sum_{j=1}^{s} Q_j(\theta)T_j(x) \right), \; s \leqslant k \qquad \ldots \quad (1.1)$$

By definition the densities (1.1) form an exponential family. Here $x \in X$, θ is a parameter taking on values in an abstract set Ω, $Q_j(\theta)$ are arbitrary functions and the functions $T_j(x)$, $h(x)$ are measurable

with respect to $\mathcal{A}$. It is convenient to introduce the "natural parameters"

$$\theta_j = Q_j(\theta), \qquad j = 1, \ldots, s, \qquad \theta = (\theta_1, \ldots, \theta_j)$$

and to rewrite (1.1) in the form

$$p_\theta(x) = C(\theta)h(x) \exp\left\{\sum_{j=1}^{s} \theta_j T_j(x)\right\}. \qquad \ldots \quad (1.2)$$

Note here that the behaviour of the parameters $\theta_1, \ldots, \theta_s$ is of great importance for the family (1.2). The simplest case is when $\theta = (\theta_1, \ldots, \theta_s)$ runs over a non-degenerate parallelopiped $\theta_i^1 \leqslant \theta_i \leqslant \theta_i^2$, $i = 1, \ldots, s$ in E_s; it is the case of a complete family.

We shall consider also the case when the parameters satisfy the relationships

$$\begin{aligned} &\Pi_1(\theta_1, \ldots, \theta_s) = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\Pi_r(\theta_1, \ldots, \theta_s) = 0, \end{aligned} \qquad \ldots \quad (1.3)$$

where $r < s$. The functions $\Pi_1, \ldots, \Pi_r$ are supposed here sufficiently smooth in a sense which we shall make more precise later.

When the parameters $\theta_1, \ldots, \theta_s$ obey the relationships (1.3) we call the family (1.2) an incomplete exponential family. There are many statistical problems for these families the investigation of which began only recently.

To begin with we consider a complete exponential family and then we shall add the relationships. Let

$$d\nu(x) = h(x)d\mu(x),$$

where μ is a new measure on (X, A) which we suppose σ-finite. We suppose the parameter space Ω consisting of all the points $\theta = (\theta_1, \ldots, \theta_s)$ for which

$$\int_{E_k} \ldots \int \exp\left\{\sum_{j=1}^{s} \theta_j T_j(x)\right\} d\nu(x) < \infty. \qquad \ldots \quad (1.4)$$

Following Lehmann [1] we call Ω the space of natural parameters; Ω is always a convex set of E_s (see [1] or [3]).

The concept of complexification of the parameter space is of great importance for application of the theory of complex variables.

Though the complex values of parameters have in general no statistical meaning the investigation of these values leads to purely statistical results. For all parts in the complexification Ω^c of Ω that is for values

$$\theta = (\theta_1+i\tau_1, \ldots, \theta_s+i\tau_s),\ (\theta_1, \ldots, \theta_s)\epsilon\Omega,\ -\infty < \tau_j < \infty,\ j = 1, \ldots, s$$

(1.4) remains true.

It is well known that the statistics $T_1(x), \ldots, T_s(x)$ in the expression (1.2) are sufficient for the parameters $(\theta_1, \ldots, \theta_s)$ and (1.2) gives the factorization of $p_\theta(x)$ in the sense of Fisher. Setting $T = (T_1(x), \ldots, T_s(x))$ we may affirm that for the family (1.2) which we denote by $\mathcal{P}_\theta$ and for any Borel set $A \subset X$ there exists a version of the conditional probability $P_\theta(A\,|\,T)$ which is independent of θ. It is interesting to discuss certain details of the concept of sufficient statistics which play a fundamental role in statistics. In general the property of a statistic T being sufficient has to mean that T "contains the whole information" about the parameter θ. So we could expect that any statistic U independent of T has to contain no information on θ and that the distribution of U is independent of θ. But it is not true in general as was shown by D. Basu in 1958 [4]. It is useful to discuss a corresponding example (though it is outside the area of exponential families).

Consider the uniform distribution on the interval $I_\theta = [\theta-\frac{1}{2}, \theta+\frac{1}{2}]$, the parameter θ taking the values $2m+\frac{1}{2}$, $m = 0, \pm1, \pm2, \ldots$.

Let $X = (x_1, x_2)$ be a sample of two independent observations from the population above. Certainly each of the observations completely determines the value of θ and is a sufficient statistic. So the statistic x_2 is independent of the sufficient statistic x_1 and at the same time x_2 contains the whole information on θ and is itself a sufficient statistic. To study this phenomenon in detail consider a measurable space $(X, \mathcal{A})$ and the family of measures $\mathcal{P} = \{p_\theta\}$ on it. Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be an abstract-valued sufficient statistic for $\mathcal{P}$ and denote by $P_\theta T^{-1}$ the measure on $\mathcal{B}$. For any $A \in \mathcal{A}$ there exists a version $P(A\,|\,T)$ of $P_\theta(A\,|\,T)$ independent of θ so that

$$P_\theta(A \cap T^{-1}B) = \int_B P(A\,|\,T)dp_\theta T^{-1},\quad B \in \mathcal{B}, \qquad \ldots\ (1.5)$$

Suppose that for all θ a set ("event") A is independent of T. Then we have

$$P_\theta(A \cap T^{-1}B) = P_\theta(A)P_\theta(T^{-1}B) = \int_B P_\theta(A)dP_\theta T^{-1}, B \in \mathcal{B} \ldots \quad (1.6)$$

By comparing (1.6) with (1.5) we conclude that

$$P_\theta(A) = P(A \mid T) \quad \ldots \quad (1.7)$$

for almost all T according to the measure $P_\theta T^{-1}$. If for the values θ_1, θ_2 of θ there exists a set $B \in \mathcal{B}$ such that $P_{\theta_1}T^{-1}(B) > 0$, $P_{\theta_2}T^{-1}(B) > 0$, then we get from (1.7)

$$P_{\theta_1}(A) = P_{\theta_2}(A).$$

Hence if it is possible to "link" any two values of θ then the set A contains no information on θ. But if the measures corresponding to different values of θ are mutually singular this conclusion may not be true and it is just the case in the example considered above. It is worthwhile to note that for the exponential families of non-degenerate distributions it is always possible to "link" any two values of the parameter.

Rather interesting results can be obtained by weakening the definition of sufficiency towards "partial sufficiency". Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a statistic and $f(x)$ a measurable vector-valued function for which the conditional expectation

$$E_\theta(f \mid T) \quad \ldots \quad (1.8)$$

does not depend upon θ. Consider all the functions $f(x)$ for which (1.8) is independent of θ. Obviously they form a linear space $\mathcal{L}$. If T is a sufficient statistic $\mathcal{L}$ coincides with the space of all f with finite expectation. In general we call such T a partially sufficient statistic with respect to $\mathcal{L}$.

The study of partially sufficient statistics began only recently and we mention here certain results obtained by A. Kagan and O. Shalaevskiĭ.

Let $x = (x_1, \ldots, x_n)$ be a sample formed by independent observations $x_1, \ldots, x_n$ with marginal density $p(x-\theta)$ (with respect to Lebesgue

measure) depending upon a location parameter θ. Consider the sample mean

$$\bar{x} = \frac{1}{n} \sum_{1}^{n} x_i$$

and the sample variance

$$s^2 = \frac{1}{n} \sum_{1}^{n} (x_i - \bar{x})^2$$

and suppose that $\bar{x}$ is partially sufficient with respect to the space $\mathcal{L} = \{cs^2,\ c \text{ real}\}$, i.e.,

$$E_\theta(s^2 \mid \bar{x}) = F(\bar{x}) \text{ is independent of } \theta. \qquad \ldots \quad (1.9)$$

The method proposed by A. Kagan and O. Shalaevskiĭ for investigation of the relation (1.9) is briefly the following. By multiplying both sides of (1.9) by $\exp it(\bar{x}-\theta)$ and then taking the expectations one can prove that $F(\bar{x})$ has to be a constant. So (1.9) is reduced to the regression-type relation

$$E_\theta(s^2 \mid \bar{x}) = \text{const.}$$

which leads, as was shown by E. Lucacs in 1942, to a differential equation for the characteristic function. The only solution of the differential equation is the characteristic function of the normal law.

The investigation of partial sufficiency of $\bar{x}$ with respect to other polynomials $P(x_1, \ldots, x_n)$ i.e., the property of $E_\theta(P(x_1, \ldots, x_n) \mid \bar{x})$ being independent of θ for independent x_i's with marginal density $p(x-\theta)$ shows that for many polynomials P this property is characteristic for the normal distribution (Yu. Linnik, A. Zinger). For non-negative x_i's with density $p\left(\frac{x}{\sigma}\right)$ depending upon a scale parameter σ the distribution of Pearson type III appears in the above situation if the density $p(x)$ satisfies certain additional assumptions but in general the investigation of the phenomenon of partial sufficiency has only begun.

REFERENCES

[1] E. L. Lehmann, *Testing Statistical Hypotheses*, N. Y., Wiley, 1959.

[2] N. Chentzov, Catagories of mathematical statistics (in Russian), *Doklady Akad. Nauk SSSR*, **164**, 3 (1965).

[3] Yu. V. Linnik, Statistical problems with nuisance parameters, Providence, *Amer. Math. Soc.*, 1968.

[4] D. Basu, On statistics independent of sufficient statistics, *Sankhyā*, **20**, pp. 3–4 (1958)

CHAPTER TWO

APPLICATIONS OF SUFFICIENT STATISTICS. INDEPENDENT STATISTICS

Sufficient statistics have interesting applications in the theory of independent statistics which we shall demonstrate now.

Recall the definition of a complete family introduced in the interesting paper by E. Lehmann and H. Scheffe [1] (see also [2]).

A family of probability distributions $\{P_\theta,\ \theta \in \Omega\}$ on a measurable space $(X, \mathcal{B})$ is said to be complete if

$$E_\theta f(x) \equiv 0,\ \theta \in \Omega \rightarrow f(x) \equiv 0 \text{ almost everywhere } P_\theta. \quad \ldots \quad (2.1)$$

A family $\{P_\theta,\ \theta \in \Omega\}$ is said to be boundedly complete if (2.1) holds for measurable bounded functions.

Revert now to exponential families (1.2). The complexification of the space Ω allows us to prove easily the following result (see [2]).

Suppose $\{P_\theta^T\}$ is the family of distributions of the sufficient statistic $T = (T_1(x), \ldots, T_s(x))$. If the parameter space Ω contains a parallelopiped of dimensions s then the family $\{P_\theta^T\}$ is complete.

Consider now a vector-valued statistic $U = U(x)$ whose distribution is independent of θ; we call such statistics similar. Similar statistics and tests connected with them are a subject of interest to us. We shall prove that any similar statistic U is independent of the complete sufficient statistic T (the pathological phenomena of Basu type discussed in the previous chapter are impossible for the exponential families).

Let $U = (U_1, \ldots, U_r)$ be a similar statistic; consider its characteristic function

$$E_\theta\{\exp i(\xi_1 U_1 + \ldots + \xi_r U_r)\} = \phi(\xi_1, \ldots, \xi_r). \quad \ldots \quad (2.2)$$

The function ϕ is independent of θ. We may rewrite (2.2) in the form

$$E_\theta\{E_\theta[\exp i(\xi_1 U_1+\ldots+\xi_r U_r)-\phi(\xi_1, \ldots, \xi_r)\,|\,T]\} = 0.$$

Since T is a sufficient statistic the expression

$$E_\theta[\exp i(\xi_1 U_1+\ldots+\xi_r U_r)-\phi(\xi_1, \ldots, \xi_r)\,|\,T] = V(\xi_1, \ldots, \xi_r, T)$$

is independent of θ. On the other hand V is a measurable (with respect to T) and bounded function with $E_\theta V \equiv 0$, $\theta \in \Omega$. Taking into account the completeness of the corresponding family we conclude that

$$E_\theta\{\exp i(\xi_1 U_1+\ldots+\xi_r U_r)\,|\,T] = \phi(\xi_1, \ldots, \xi_r)$$

which proves independence U of T.

This result has a number of useful applications (see [3, 4]) of which we give here four examples.

1. Suppose that $x_1, \ldots, x_n$ are independent normal $\mathcal{N}(a, \sigma^2)$ observations, σ^2 being known, $a \in (-\infty, \infty)$ being an unknown parameter. Let $g(x_i - x_j)$ be a vector-valued Borel measurable function depending only upon differences of the observations. It is similar relative to a and hence independent of the sufficient statistic $\bar{x}$.

2. In the preceding example let $\theta = (a, \sigma^2)$ be the unknown parameter. Then the statistic $(\bar{x}, s^2)$ is sufficient for θ. Consider a vector-valued measurable function $g\left(\frac{x_\alpha - x_\beta}{x_\gamma - x_\delta}\right)$ invariant with respect to affine transformations $x_i \to cx_i + d$. It is a similar statistic and hence is independent of the pair $(\bar{x}, s^2)$. Thus, for example, the statistic

$$\frac{m_\nu}{m_2^{\nu/2}} = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^\nu$$

for all integers ν is independent of $(\bar{x}, s^2)$. This well known result is proved in the book by Cramer with the help of direct and rather hard calculations.

3. Let $x_1, \ldots, x_n$ be independent observations distributed according to Pearson type III density of the form

$$p(x, \gamma) = \begin{cases} \dfrac{\gamma^m}{\Gamma(\gamma)}\, x^{m-1} \exp(-\gamma x), & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

with fixed m and $\gamma \in (0, \infty)$ as a parameter. $\bar{x}$ is a sufficient statistic for γ. All vector-valued measurable functions invariant with respect to homothetic transformations are independent of $\bar{x}$. For example, the random vector

$$\left(\frac{x_1}{x_1+\ldots+x_n}, \frac{x_2}{x_1+\ldots+x_n}, \frac{x_{n-1}}{x_1+\ldots+x_n}\right)$$

is independent of $x_1+\ldots+x_n$.

4. Consider now independent vectors $(x_1, y_1), \ldots, (x_n, y_n)$ each following two-dimensional normal distribution with the density

$$p_\theta(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp -\frac{1}{2}\left[\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right],$$

where $\theta = (a_1, a_2, \sigma_1^2, \sigma_2^2)$. Here $a_i \in (-\infty, \infty)$, $\sigma_i^2 \in (0, \infty)$ and the distributions of sufficient statistics

$$\bar{x}, \bar{y}, s_1^2 = \frac{1}{n}\sum_1^n (x_i-\bar{x})^2, \; s_2^2 = \frac{1}{n}\sum_1^n (y_i-\bar{y})^2$$

from a complete exponential family.

Let us take a vector-valued statistic $g(x_1, \ldots, x_n; y_1, \ldots, y_n)$ invariant with respect to affine transformations $x_i \to \alpha x_i+\beta$, $y_i \to \gamma y+\delta$. With the help of one of these transformations the density $p_\theta(x, y)$ can be always reduced to the density with $a_1 = a_2 = 0$, $\sigma_1 = \sigma_2 = 1$, hence the distribution of g does not depend upon the parameters $a_1, a_2, \sigma_1^2, \sigma_2^2$ and g itself is independent of $(\bar{x}, \bar{y}, s_1^2, s_2^2)$. In particular the sample correlation coefficient

$$r = \frac{\frac{1}{n}\sum_1^n (x_i-\bar{x})(y_i-\bar{y})}{s_1 s_2}$$

is independent of $(\bar{x}, \bar{y}, s_1^2, s_2^2)$. We shall use this fact later.

Other interesting examples can be found in the note by T. Feigel'son [4].

We pass now to an application of the famous theory of analytical sheaves by H. Cartan to the theory of independent statistics which

allows us to elucidate certain interesting properties of independent statistics.

Let (X, Y) be a two-dimensional random vector. The independence of its components X and Y is equivalent to the relation

$$E \exp i(t_1X+t_2Y) = E \exp it_1X. E \exp it_2Y \quad \ldots \quad (2.3)$$

holding for all values of t_1 and t_2. If (X', Y') is a vector with independent components distributed as X and Y respectively we shall denote this fact by writing

$$X \cong X', \quad Y \cong Y'.$$

Then for all real t_1, t_2 (2.3) may be rewritten as

$$E \exp i(t_1X+t_2Y) = E \exp i(t_1X'+t_2Y'),$$

i.e.,

$$t_1X+t_2Y \cong t_1X'+t_2Y'$$

for all real t_1, t_2. Thus the independence of X and Y is equivalent to the identical distribution of t_1X+t_2Y and $t_1X'+t_2Y'$ and the problem of independence of statistics is connected with the problem of identical distribution.

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space equipped with a probability measure μ and $\{Q(X, a_1, \ldots, a_s)\}$ be a family of statistics depending on real parameters $a_1, \ldots, a_s$ which vary in a closed bounded parallelopiped Ω. We get

$$(a_1, \ldots, a) = a, \quad Q(X, a_1, \ldots, a_s) = Q(X, a).$$

If two statistics $Q(X, a)$ and $Q(X, a')$, a, $a' \in \Omega$, are identically distributed, i.e. if

$$Q(X, a) \cong Q(X, a') \quad \ldots \quad (2.4)$$

and if all the moments $E[Q(X, a)]^r$, $E[Q(X, a')]^r$, $r = 0, 1, 2, \ldots$, exist then the relation (2.4) implies that

$$E[Q(X, a]^r = E[Q(X, a')]^r, \quad r = 0, 1, 2, \ldots. \quad \ldots \quad (2.5)$$

Conversely, if the problem of moments were determinate in the present case, then (2.4) would follow from (2.5). The sufficient condi-

tions for the problem of moments to be determinate are well-known (see [5]) :

$$\sum_{r=1}^{\infty} \alpha_{2r}^{-1/2r} = \infty,$$

where α_{2r} is the moment of order $2r$. Though this condition is only sufficient it is almost necessary. But it is not satisfied even in the case when X is normally distributed and Q is a polynomial statistic. That is why we shall consider the generalizations of the moments (2.5) which we shall call ϕ-moments. We introduce, a family of continuous functions $\{(\phi_r U),\ r = 1,\ 2, \ldots\}$ of a real argument U. Suppose that the family $\{\phi_r(U)\}$ has the property (D) :

(D) Determinancy of the analogue of the problem of moments : the mathematical expectations (ϕ-moments)

$$E\phi_r(Q(X, a)), \quad a \in \Omega, \qquad \ldots \quad (2.6)$$

exist and their value determine the distribution of $Q(X, a)$ up to μ-measure zero.

Furthermore, we suppose that the statistics $Q(X, a)$ possess property (Γ) relative to the family $\{\phi\ (U)\}$: (Γ) Holomorphic connectedness. Let us define

$$E[\phi_r(Q(X, a))] = \phi_r(a_1, \ldots, a_s). \qquad \ldots \quad (2.7)$$

Then the region of s complex variables $a_1, \ldots, a_s$ has to contain a bounded parallelopiped Ω_1 such that $\Omega_1 \supset \operatorname{Int} \Omega_1 \supset \Omega$, on which all the functions (2.7), $r = 1, 2, \ldots$, exist and are holomorphic.

According to the property (D) if for $a \in \Omega$ and $a' \in \Omega$

$$E\phi_r[Q(X, a)] = E\phi_r[Q(X, a')], \ r = 1, 2, \ldots, \qquad \ldots \quad (2.8)$$

then

$$Q(X, a) \cong Q(X, a') \qquad \ldots \quad (2.9)$$

(and conversely). Now the condition (Γ) of holomorphic connectedness leads to the curious fact that the property (2.9) of $Q(X, a)$ and $Q(X, a')$ being identically distributed follows from a finite number of relations (2.8) between the ϕ-moments.

Theorem 2.1. *Suppose that for a given space* (X, A, μ) *and families* $\{Q(X, a)\}$ *and* $\{\phi_r\}$ *the conditions* (D) *and* (Γ) *are satisfied. Then there exists a finite constant* R_0 *such that the relations*

$$E\phi_r[Q(X, a)] = E\phi_r[Q(X, a')], \quad r = 1, \ldots, R_0; \quad a, a' \in \Omega$$

are equivalent to

$$Q(X, a) \cong Q(X, a').$$

To prove Theorem 2.1 we note that (2.4) follows from the countable set of the relations (2.5) which may be put in the form

$$\phi_r(a_1, \ldots, a_s) - \phi_r(a_1', \ldots, a_s') = 0, \quad r = 1, 2, \ldots. \qquad \ldots \quad (2.10)$$

Consider the parallelopiped $\Omega_1 \times \Omega_1$ formed by the points (a, a') and construct the ideal $\mathcal{I}$ generated by the differences

$$\phi_r(a_1, \ldots, a_s) - \phi_r(a_1', \ldots, a_s'), \quad r = 1, 2, \ldots,$$

within the ring of holomorphic functions on $\Omega_1 \times \Omega_1$. According to the well known theorem by H. Cartan in the theory of analytical sheaves the ideal $\mathcal{I}$ has a finite basis and it is generated by a finite number of differences

$$\phi_r(a_1, \ldots, a_s) - \phi(a_r, \ldots, a_s'), \quad r = 1, \ldots, R_0.$$

Thus if the relations (2.10) are satisfied for $r \leqslant R_0$, they are also satisfied for $r > R_0$ which proves Theorem 2.1.

An analogous assertion can be made for the problem of independence of two statistics. Consider the same measurable space $(\mathcal{X}, \mathcal{A})$ equipped with a measure μ and two families of statistics $\{Q(X, a); a \in \Omega\}$ and $\{T(X, a'); a \in \Omega'\}$, Ω and Ω' being parallelopipeds of the same kind as Ω from Theorem 2.1. Let us introduce two families $\{\phi_r(U), r = 1, 2, \ldots\}$ and $\{\psi_t(U), t = 1, 2, \ldots\}$ of continuous functions such that the functions

$$E[\phi_r(Q(X, a))\psi_t(T(X, a'))] = \phi_{r,t}(a, a'), \quad r, t = 1, 2, \ldots,$$

and

$$E\phi_r[Q(X, a)] = \chi_r(a), \; E\psi_t[T(X, a')] = \tau_t(a'), \quad r, t = 1, 2, \ldots,$$

exist in $\Omega \times \Omega'$.

Furthermore suppose that the following condition (D') is satisfied : $(\mathfrak{D}')$:

The countable set of the relations

$$\phi_{r,t}(a, a') = \chi_r(a)\tau_t(a'), \quad r, t = 1, 2, \ldots,$$

holding for all the points $a \in \Omega$, $a' \in \Omega'$ implies independence of the statistics $Q(X, a)$ and $T(X, a')$. Formulate now a condition of "holomorphic connectedness".

(Γ') There exist two closed parallelopipeds Ω_1 and Ω_1' such that $\Omega_1 \supset \text{Int } \Omega_1 \supset \Omega$, $\Omega_1' \supset \text{Int } \Omega_1' \supset \Omega'$ and the functions $\chi_r(a)$, $\tau_t(a')$, $\phi_{r,t}(a, a')$, $r, t = 1, 2, \ldots$ are holomorphic for $a \in \Omega_1$, $a' \in \Omega_1'$, $(a, a') \in \Omega_1 \times \Omega_1'$ respectively.

Theorem 2.2. *Suppose that for given families* $\{Q(X, a)\}$, $\{T(X, a')\}$, $\{\phi_r\}$ *and* $\{\psi_t\}$ *the conditions* (D') *and* (Γ') *are satisfied. Then there exists a constant* R_1 *such that the relations*

$$E\phi_r(Q(X, a))\psi_t(T(X, a')) = E\phi_r(Q(X, a))\, E\psi_t(T(X, a')),$$
$$r, t = 1, 2, \ldots, R_1 \quad \ldots \quad (2.11)$$

are equivalent to independence of the statistics $Q(X, a)$ *and* $T(X, a')$.

The proof is analogous to that of Theorem 2.1. We consider the ideal I generated by the differences

$$\phi_{r,t}(a, a') - \chi_r(a)\tau_t(a'). \quad \ldots \quad (2.12)$$

According to the above mentioned theorem by H. Cartan the ideal I has a finite basis generated by the functions (2.12) with $r, t = 1, 2, \ldots, R_1$, R_1 being a finite constant. So if the relations (2.11) are satisfied for $r \leqslant R_1$, $t \leqslant R_1$ they are also satisfied for all values of r and t which implies independence of $Q(X, a)$ and $T(X, a')$. We see that independence of two statistics in a rather general situation is equivalent to a finite number of uncorrelatedness relations. The constant R_1 is determined by the families under consideration. It would be interesting to study this "constant of uncorrelatedness leading to independence" and its behaviour for different families of distributions.

Let us give now two examples.

1. Let $\{Q(X, a)\}$ be a family of positive polynomials, for example, the squares of certain other polynomials; $a \in M$ where $M \subset E_s$ is a closed parallelopiped. Let $\{\phi_r(u)\}$ be the set of functions

$$\phi_r(u) = u^{r-1} \exp(-\sigma(u)), \quad r = 1, 2, \ldots$$

$\sigma(u)$ being a non-negative polynomial of degree $\geqslant 2$ for $u > 0$. Theorem 2.1 may be applied in this situation.

2. Let us set $\psi_t(u) = u^{t-1} \exp(-\sigma_1(u))$, where $\sigma_1(u)$ is either the same as $\sigma(u)$ or another non-negative polynomial of degree $\geqslant 2$. If $\{Q(X, a)\}$, $\{T(X, a')\}$ are two families of positive polynomials, Theorem 2.2 may be applied.

REFERENCES

[1] E. L. Lehmann and H. Scheffe, Completeness, similar regions and unbiased estimates. *Sankhyā*, **10**, 1950, 305–340.

[2] E. L. Lehmann, *Testing Statistical Hypothesis*, N.Y., Wiley, 1959.

[3] R. Hogg and A. Craig, Sufficient statistics in elementary distribution theory, *Sankhyā*, **17**, 1956, 3.

[4] T. Feigel'son, On a simple method of establishing independence of statistics (in Russian), *Vestnik Leningrad. Univ., Ser. Matem., Mechan., Astron.*, **19**, 3 (1964), 157–158.

[5] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, N.Y., 1943.

CHAPTER THREE

ON "UNLINKING" INDEPENDENT STATISTICS

In many problems of multivariate analysis particularly in construction of tests one has often recourse to independent statistics. In all known tests these statistics are rational functions of observations. This leads to the folowing problem.

Suppose $x_1, \ldots, x_n$ are independent normal $\mathcal{N}(a, \sigma^2)$ observations. How to characterise all vectors $(r_1, \ldots, r_k)$ whose components are independent rational statistics ? We may always assume without loss of generality that $a = 0$, $\sigma = 1$. The example 3 of the previous chapter gives such a set

$$\left(\frac{x_1^2 + \ldots + x_r^2}{x_1^2 + \ldots + x_n^2} \, , \; x_1^2 + \ldots + x_r^2 \right). \qquad \ldots \quad (3.1)$$

But we have no means to describe the phenomenon of independence of rational statistics systematically.

We are now in a position to formulate the following conjecture in this direction.

Conjecture. If two polynomials $Q_1(x_1, \ldots, x_n)$ and $Q_2(x_1, \ldots, x_n)$ of independent normal $\mathcal{N}(0, 1)$ variables are themselves independent they can be always "unlinked" : i.e., there exists an orthogonal transformation

$$F : (x_1, \ldots, x_n) \to (y_1, \ldots, y_n)$$

such that the polynomials $Q_1 = Q'_1(y_1, \ldots, y_n)$, $Q_2 = Q'_2(y_1, \ldots, y_n)$ depend upon disjoint subsets of new varibles $y_1, \ldots, y_n$ and are functionally independent.

It is probable that in this conjecture polynomials may be replaced by holomorphic functions. But as we have seen in the example (3.1) the singularities at a finite distance from the origin in one of the functions makes our conjecture incorrect.

Till now the conjecture is neither proved nor disproved. There exists an approach to it based on algebraic geometry (see [1]). It is possible to show that for fixed n (number of variables) and fixed degrees of both polynomials the conjecture can be verified in a finite number of calculations. But the calculations are extremely complicated and do not allow one to obtain the solution in the general situation.

In case of $n = 2$ the conjecture is true. If at least one of the polynomials is symmetric it is possible to obtain certain other results. Note, by the way, that in case of independent identically distributed observations symmetric statistics are widely used because of natural arguments. We shall summarize here the results obtained when at least one of the polynomials is symmetric.

Let $Q_1(x_1, \ldots, x_n)$ be a non-constant polynomial and $Q_2(x_1, \ldots, x_n)$ be a symmetric homogeneous polynomial of degree m. Suppose that for independent normal $\mathcal{N}(0, 1)$ variables $x_1, \ldots, x_n$ these polynomials are independent. We may always assume Q_1 and Q_2 non-negative because otherwise we could consider Q_1^2 and Q_2^2 instead of Q_1 and Q_2. Consider now the surface

$$Q_2(x_1, \ldots, x_n) = 1. \qquad \ldots \quad (3.2)$$

Since the polynomial Q_2 may be degenerate, i.e., reducible by a linear transformation to a polynomial depending on $k < n$ variables, the surface (3.2) may be a cylinder. Let $k < n$ be the "true number" of variables of the polynomial $Q_2(x_1, \ldots, x_n)$. Construct now the sphere

$$x_1^2 + \ldots + x_n^2 = \epsilon^2, \qquad \ldots \quad (3.3)$$

ϵ being small enough for the sphere (3.3) to be situated "within" the surface (3.2). Let us stretch the sphere (3.3) till it becomes tangent with the surface (3.2). Let $(\xi_1, \ldots, \xi_n)$ be a point at which the surfaces are tangent. Since the surface (3.3) is also symmetric there are $n!$ points of tangency obtained by $n!$ permutations of the coordinates $(\xi_1, \ldots, \xi_n)$. These points may all coincide : $\xi_1 = \ldots = \xi_n$; since Q_2 may depend in fact on $k < n$ variables there may exist linear constraints between the points of tangency (and they always exist because we have $n!$ points in the n-dimensional space). There may

exist degeneracy of another type : the sphere may tangent with the surface (3.2) along a curve or along a surface itself. But these will be, so to say, "rare" or "pathological" cases. In general the $n!$ points of tangency are isolated and there are k linearly independent points among them. This is the general case. Observe that if $k = n$ (the polynomial Q_2 is non-degenerate) the condition of existence of $k = n$ linearly independent points of tangency is very simple : the point $(\xi_1, \ldots, \xi_n)$ must not belong either to the line $\xi_1 = \ldots = \xi_n$ or to the plane $\xi_1 + \ldots + \xi_n = 0$.

So we suppose that there exist k linearly independent points of tangency :

$$M_1 = (\xi_1, \ldots, \xi_{1,n})$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$M_k = (\xi_{k,1}, \ldots, \xi_{k,n}).$$

Let us make an orthogonal transformation $F : (x_1, \ldots, x_n) \to (y_1, \ldots, y_n)$ which transforms the polynomial $Q_2(x_1, \ldots, x_n)$ into a non-degenerate polynomial $Q_2'(y_1, \ldots, y_k)$. The polynomial $Q_1(x_1, \ldots, x_n)$ will be transformed into a polynomial $Q_1'(y_{\alpha_1}, \ldots, y_{\alpha_s})$ which also may be assumed non-degenerate. The transformation F leaves the sphere (3.3) invariant and we have the new tangency points $M_2', \ldots, M_k'$ also linearly independent. Let us take the point M_1' and consider the second orthogonal transformation

$$F' : (y_1, \ldots, y_k) \to (y_1', \ldots, y_k')$$

transferring the point M_1' on the subspace $y_2' = y_3' = \ldots = y_k' = 0$ (the other varables are arbitrary). Let us take a sufficiently large number R and stretch the surface (3.1) and the tangent sphere by constructing the sphere

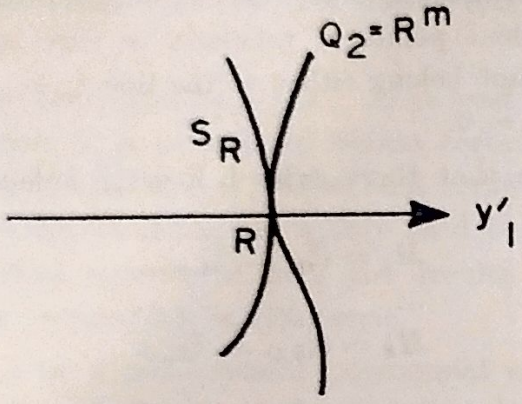
$$S_R : y_1'^2 + \ldots + y_n'^2 = R^2 \qquad \ldots \quad (3.4)$$

and the surface

$$Q_2(y_1', \ldots, y_k') = R^m. \qquad \ldots \quad (3.5)$$

At the points of tangency between (3.4) and (3.5) we have no singularities (angles etc.) because the derivatives $\frac{\partial Q_1}{\partial y}$ may equal zero simultaneously only at the origin according to Euler's theorem on

homogeneous functions. Consider now the event $Q_2 = R^m$ and the induced conditional distribution of $(y_1', \ldots, y_n')$. The original (unconditional) distribution of $(y_1', \ldots, y_n')$ is the same as that of $(x_1, \ldots, x_n)$ because of orthogonality of the transformations. Since R is large the distribution is concentrated at a neighbourhood of the



tangency points, i.e., the conditional distribution will be concentrated at a neighbourhood of the size 0(1) of the tangency points. If we construct a cone with the vertex at the origin, the base encloses the tangency points and the vertex angle is $\frac{K}{R}$, R being large enough, the conditional probability of the outside part of the cone given (3.5) will be of the order $\exp(-cK^2)$ and for sufficiently large K it is negligible.

Reverting to the first polynomial $Q_1(y_1', \ldots, y_n')$ we are going to prove that it is independent of y_1'. If it is not the case then

$$Q_1 = P_0(y_2', \ldots, y_n')y_1'^r + P_1(y_2', \ldots, y_n')y_1'^{r-1} + \ldots + P_r(y_2', \ldots, y_n') \quad \ldots (3.6)$$

where the polynomials $P_0, P_1, \ldots, P_r$ may be degenerate.

We shall show firstly that $P_0(y_2', \ldots, y_n') \equiv 0$. If it is not the case then with conditional probability $\geqslant \alpha_0$, where α_0 is a certain constant independent of R we would have $c_1 \leqslant y_i' \leqslant c_2$, $i = 2, \ldots, n$, c_1, c_2 being arbitrary constants such that $|c_i| \leqslant K$. To the whole set of values $(y_2', \ldots, y_n')$ there corresponds a point y_1' with $y_2' - R = 0(1)$ on the surface (3.5), the probability of this being $\geqslant \alpha_0$. But then it is possibe to choose c_1 and c_2 in such a way that $|P_0(y_2', \ldots, y_n')| > c_3$ and

$$|Q_1| > \frac{1}{2}c_3R^2 \quad \ldots (3.7)$$

with conditional probability $\geqslant \alpha_0$. This is impossible if Q_1 is independent of Q_2 because the probability of (3.7) is equal to $0(\exp -c_4 R^2)$, $R \to \infty$. Hence $P_0 \equiv 0$. Now we take the next member and work in the same way getting as a result that $Q_1 = P_r(y_2', \ldots, y_n')$ does not depend on y_1'.

Then we may take another point of tangency between (3.4) and (3.5) and proceed in the same way. Likewise we can prove the following lemma.

If the non-degenerate polynomial $Q_2(y_1, \ldots, y_k)$ depends on variables $y_1, \ldots, y_k$, then the polynomial $Q_1(y_1, \ldots, y_n)$ is functionally independent of k linear forms

$$L_1 = L_1(y_1, \ldots, y_k)$$
$$\ldots \quad \ldots \quad \ldots$$
$$L_k = L_k(y_1, \ldots, y_k).$$

If the tangency points are linearly independent the forms $L_1, \ldots, L_k$ are so as well.

But then Q_1 is functionally independent of $y_1, \ldots, y_k$ which proves "unlinking" of Q_1 and Q_2. Observe that instead of the tangency points we could consider any points sufficiently close to the tangency points without any change in our arguments. This closeness determined by the constant R must depend on Q_1. Thus our conjecture is true in the case when Q_2 is sufficiently close to a symmetric polynomial.

In the case of $n = 2$ the conjecture is true for arbitrary independent polynomials because one point of tangency must always exist and our arguments can be carried out easily.

It would be interesting to carry through the investigation of the conjecture and to study the rational statistics for which the conjecture is false.

REFERENCE

[1] Yu. V. Linnik, Sur certaines questions de la statistique analytique (in French), *Ann Fac. Sci. Univ. Clermont-Ferrand, Math.* N 8, 1962, 53–61.

CHAPTER FOUR

PROBLEMS WITH NUISANCE PARAMETERS. VERIFIABLE FUNCTIONS. SIMILAR TESTS

Let (X, A) be a measurable space equipped with a family of probability measures $\{P_\theta\}$ dominated by a σ-finite measure μ so that there exist densities $\frac{dP_\theta}{d\mu} = p_\theta(x)$. We assume that the parameter $\theta = (\theta_1, \ldots, \theta_s)$ varies in a close parallelopiped $\Omega \subset E_s$.

Furthermore we suppose that for all x the density $p_\theta(x) = p(x; \theta_1, \ldots, \theta_s)$ is sufficiently smooth with respect to θ in a sense which we make precise later. Instead of the parameters $\theta_1, \ldots, \theta_s$ it is possible to introduce other parameters $\gamma_1(\theta_1, \ldots, \theta_s) \ldots, \gamma_s(\theta_1, \ldots, \theta_s)$ where the functions $\gamma_1, \ldots, \gamma_s$ are chosen in a convenient way preserving the regularity properties of $p_\theta(x)$.

Suppose we have to test a hypothesis H_0 concerning the parameters $\theta_1, \ldots, \theta_q$ $(q < s)$, H_0 may have the form

$$(\theta_1, \ldots, \theta_q) \in \omega$$

where ω is a subset of the set of possible values of $\theta_1, \ldots, \theta_q$. In H_0 we make no assumptions about the other parameters $\theta_{q+1}, \ldots, \theta_s$ but since they figure in $p_\theta(x)$ it is necessary to bear them in mind. Following Hotelling we call $\theta_{q+1}, \ldots, \theta_s$ the nuisance parameters for the given problem.

Often the hypothesis H_0 concerns the behaviour of certain functions $\gamma_1(\theta_1, \ldots, \theta_s), \ldots, \gamma_q(\theta_1, \ldots, \theta_s)$. If the functions $\gamma_1, \ldots, \gamma_q$ are functionally independent and sufficiently smooth it is possible to introduce $(s-q)$ functions $\gamma_{q+1}(\theta_1, \ldots, \theta_s), \ldots, \gamma_s(\theta_1, \ldots, \theta_s)$ which form together with $\gamma_1, \ldots, \gamma_q$ a new system of parameters with the same properties as the initial one. Then for the problem of testing H_0 it is natural

to consider $\gamma_{q+1}(\theta_1, \ldots, \theta_s), \ldots, \gamma_s(\theta_1, \ldots, \theta_s)$ as nuisance parameters. We consider a number of examples.

1. Test of a linear hypothesis in the scalar case (see, for example [1]). Let $x_1, \ldots, x_n$ be independent normally distributed variables with mean values $\xi_1, \ldots, \xi_n$ and common variance σ^2. It is known that the vector of mean values $(\xi_1, \ldots, \xi_n)$ belongs to a linear t-dimensional space $\Pi_\Omega \subset E_n$. The hypothesis H_0 to be tested is that the vector $(\xi_1, \ldots, \xi_n)$ belongs to a space $\Pi_\omega \subset \Pi_\Omega$ of dimension $t-r < t$.

Here the parameters are $(\theta_1, \ldots, \theta_s)$, $s = t+1$, where $\theta_1, \ldots, \theta_t$ denote the coordinates of Π_Ω and $\theta_{t+1} = \sigma$.

The hypothesis H_0 introduces r relationships between the first t parameters and we have $r+1$ nuisance parameters.

For example in the scheme of linear regression $\xi_j = \alpha+\beta t_j$, $j = 1, \ldots, n$ when t_j's are known and the hypothesis to be tested is $H_0 : \beta = 0$ the space Π_Ω consists of vectors $\alpha(1, \ldots, 1)+\beta(t_1, \ldots, t_n)$. We have in this situation $t = 2$, $s = 2+1 = 3$ and the hypothesis H_0 leaves two nuisance parameters α and σ.

It is possible to consider also a linear non-homogeneous hypothesis. For example the hypothesis $H'_0 : \alpha = \alpha_0$, $\beta = \beta_0$ fixing the values of the regression coefficients can be reduced to the hypothesis H''_0 : $\alpha = \beta = 0$ by introducing the new variables $y_j = \alpha_j-\alpha_0-\beta_0 t_j$. We have in this situation one nuisance parameter σ.

2. A case of particular importance is that of Student's problem : test the hypothesis $H_0 : a = a_0$ on the basis of independent observations $x_1, \ldots, x_n$ following normal distribution $N(a, \sigma^2)$. Here σ is a nuisance parameter.

3. For the same $x_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$ we have to test the hypothesis $\gamma(a, \sigma) = \gamma_0$ where $\gamma(a, \sigma)$ is a function with reasonable behaviour. Here it is possible to find another function $\gamma_1(a, \sigma)$ and to change the parameters : $(a, \sigma) \to (\gamma(a, \sigma), \gamma_1(a, \sigma))$. Then $\gamma_1(a, \sigma)$ will be a nuisance parameter.

4. The Behrens-Fisher problem. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be two samples drawn from normal populations $\mathcal{N}(a_1, \sigma_1^2)$ and $\mathcal{N}(a_2, \sigma_2^2)$ respectively. We want to test the hypothesis $H_0 : a_1-a_2 = 0$. It is possible to consider a_1, σ_1, σ_2 as nuisance parameters. This

problem, which is of keen interest, is a special case of the generalized linear hypothesis when not only the mean values but also the variances change from one sample to the other.

This problem is one of the most wellknown problems of the theory of testing hypotheses in the presence of nuisance parameters and we shall pay special attention to it.

5. In the situation of Example 4 let us consider the problem of testing the hypothesis $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = \gamma_0$. Here the nuisance parameters are, for example, a_1, a_2 and σ_2.

Let us revert now to the measurable space $(X, \mathcal{A})$ and the family of densities $\{p(x; \theta_1, \ldots, \theta_s)\}$ introduced earlier. Consider a hypothesis H_0 of the form : $\theta_1 = \theta_1^{(0)}, \ldots, \theta_q = \theta_q^{(0)}$, $q < s$. As it has already been noted, the hypothesis H_0' that there are q relationships of the form $\gamma_1(\theta_1, \ldots, \theta_s) = 0, \ldots, \gamma_q(\theta_1, \ldots, \theta_s) = 0$ among the parameters when certain analytic conditions are imposed on $\gamma_1, \ldots, \gamma_q$ can be reduced to H_0. An alternative to H_0 is the hypothesis that the parameters $\theta_1, \ldots, \theta_s$ belong to the complement of the set defined by the hypothesis H_0.

By special case H_0' of the hypothesis H_0 we shall mean hypotheses which attach to the parameters $\theta_1, \ldots, \theta_s$ any values consistent with H_0.

The hypothesis H_0 and its special cases are in general composite (not simple) statistical hypotheses in the sense that they do not fix the values of nuisance parameters and hence do not specify the density $p(x, \theta_1, \ldots, \theta_s)$.

We shall test the hypothesis H_0 with the aid of tests $\phi(x)$, $x \in X$, randomized (see [1]) in general; they are non-randomized when $\phi(x)$ assumes only the values 0 and I. The meaning of $\phi(x)$ is the following : it is the probability of rejecting H_0 when the value of the observation equals x. The mathematical expectation

$$E_\theta \phi(x) = \phi(\theta_1, \ldots, \theta_s) \qquad \ldots \quad (4.1)$$

is the power function of the test. Its values for those $\theta_1, \ldots, \theta_s$ which agree with H_0, i.e.

$$E_\theta \phi(x) = \phi(\theta_1^{(0)}, \ldots, \theta_q^{(0)}, \theta_{q+1}, \ldots, \theta_s) = \alpha(\theta_{q+1}, \ldots, \theta_s) \quad \ldots \quad (4.2)$$

are called the level of the test whereas its values for those $\theta_1, \ldots, \theta_s$ which agree with the alternative H_1 are called the power of the test :

$$E_\theta \phi(x) = \beta(\theta_1, \ldots, \theta_s), \ (\theta_1, \ldots, \theta_s) \in H_1. \qquad \ldots \quad (4.3)$$

It is natural to pose the following question : whether or not it is possible to test the hypothesis H_0 with the aid of a test whose power function is independent of the nuisance parameters $\theta_{q+1}, \ldots, \theta_s$. If the answer is affirmative then it would be interesting to compare the quality of these tests with the quality of tests constructed in the situation of known $\theta_{q+1}, \ldots, \theta_s$.

If we do not know the values of nuisance parameters then utilization of the tests whose power function is indepencent of nuisance parameters (we shall call them tests with invariant power function) is quite natural. For other tests we simply cannot calculate their power functions and hence cannot estimate their quality. There exists always a (trivial) test with invariant power function : the test $\phi(x) \equiv \alpha$ where the level α is given in advance. One can easily see that this test is completely useless. The thing of real interest is the construction of tests whose power function is independent of nuisance parameters but does depend on the basic parameters.

Supposing again that the hypothesis H_0 is of the form : $\gamma_1(\theta_1, \ldots, \theta_s) = \gamma_{10}, \ldots, \gamma_q(\theta_1, \ldots, \theta_s) = \gamma_{q0}$, we can give the following definition. The functions $\gamma_1, \ldots, \gamma_q$ $(q < s)$ are called verifiable if there exists a nontrivial test $\phi(x)$ $(\phi \not\equiv \text{const})$ such that its power function $\phi(\theta_1, \ldots, \theta_s)$ depends only on $\gamma_1, \ldots, \gamma_q$:

$$E_\theta \phi(x) = F(\gamma_1, \ldots, \gamma_q). \qquad \ldots \quad (4.4)$$

If $x_1, x_2, \ldots$ are independent identically distributed random variables it is possible also to give the definition of sequentially verifiable functions. In this scheme of sequential analysis due to A. Wald the separate observations $x_1, x_2, \ldots$ are obtained one after another and the decision to stop the process of observation depends on the observations already obtained. Sequential verifiability of functions $\gamma_1, \ldots, \gamma_q$ means that there exists a sequential test ϕ of the hypothesis

$$\gamma_1(\theta, \ldots, \theta_s) = \gamma_{10}, \ldots, \gamma_q(\theta_1, \ldots, \theta_s) = \gamma_{q0}$$

with finite average number of observations such that the power function of ϕ depends only on $\gamma_1, \ldots, \gamma_q$.

So we formulate the following natural problem. For a given family of densities $p(x; \theta_1, \ldots, \theta_s)$ to characterize the functions $\{\gamma_1(\theta_1, \ldots, \theta_s), \ldots, \gamma_q(\theta_1, \ldots, \theta_s)\}$, $q < s$ verifiable on the basis of independent observations $x_1, x_2, \ldots$ with the marginal density $p(x; \theta)$. We wish to describe first the functions which are verifiable in the scheme of fixed-size sample and then the ones which are sequentially verifiable. The investigation of this problem began only recently [2, 3, 4, 5].

Only the (non-sequential) case of normal observations $x_1, \ldots, x_n$, $x_i \sim \mathcal{N}(a, \sigma^2)$ and one function $\gamma(a, \sigma)$ has been investigated rather completely. The most general result here is due to V. Palamodov [5].

It is easy to see that the property of being verifiable for the functions $\{\gamma_1(\theta_1, \ldots, \theta_s), \ldots, \gamma_q(\theta_1, \ldots, \theta_s)\}$ is invariant with respect to one-to-one transformations of the parameters. In case of normal sample mentioned above the function $\gamma(a, \sigma) \equiv \sigma$ is verifiable : it is possible, for example, to take the non-randomized test

$$\phi = \begin{cases} 1, & \text{if } s^2 > c, \\ 0, & \text{if } s^2 \leqslant c, \end{cases}$$

where s^2 is sample variance. Since the statistic s^2 is similar with respect to parameter a, the power function of ϕ depends only on σ. Hence any monotone function of σ is non-sequentially verifiable. So it is interesting to consider functions $\gamma(a, \sigma)$ which depend not only on σ. V. Palamodov [5] succeeded in classifying the verifiable polynomials $p\left(a, \frac{1}{\sigma^2}\right)$ which do not depend only on σ. His theorem is the following.

Theorem 4.1. *Among the polynomials* $p\left(a, \frac{1}{\sigma^2}\right)$ *not depending only on* $\frac{1}{\sigma^2}$ *the only non-sequentially verifiable* polynomials are*

$$\frac{1}{\sigma^2}(a^2 + Aa + B) \text{ with } A^2 - 4B \leqslant 0.$$

* The definition of verifiability is here modified a little; see chapter eleven.

From here it follows at once that the function $\gamma(a, \sigma) = a$ is non-verifiable (a result in this direction is contained in a paper by G. Dantzig [2]) whereas the function $\frac{a}{\sigma}$ is verifiable. According to the famous result by C. Stein [6] the non-verifiable function $\gamma = a$ is sequentially verifiable by making use of two samples drawn one after another. This is the only result on sequential verifiability known to the author. Finally we mention a result (see [3, 4]) concerning the functions $\gamma = \frac{a}{\sigma^p}$; they are sequentially verifiable only if $p = 1$.

We shall state later the results obtained up to date including the method due to V. Palamodov. Here we have mentioned them to approach in a natural way the concept of similar tests. We have already introduced the concept of a similar statistic : this is a statistic whose distribution does not depend upon the parameters $\theta_1, \ldots, \theta_s$. We can generalize this definition by calling a statistic $\mathcal{U}$ partially similar if

$$E_\theta \mathcal{U} \text{ is independent of } \theta.$$

This definition brings together the theory of similar tests from one side and the theory of unbiased estimators from the other. Another approach to the theory of similar tests is connected with the notion of unbiased tests.

Consider as usual a space $(X, \mathcal{A})$ with a family of probability measures $\{P_\theta; \theta \in \Omega\}$ on it and two (composite, in general) hypotheses $H_0 : \theta \in \omega \subset \Omega$ and $H_1 : \theta \in \Omega \setminus \omega$. We test the hypothesis H_0 with the aid of a test ϕ (randomized, in general) whose power function is $E_\theta \phi = \phi_\phi(\theta)$.

A test ϕ is said to be unbiased if the probability of rejecting H_0 when it is true is not less than the probability of rejecting H_0 when it is false. Letting $\alpha = \sup_{\theta \in \omega} \phi_\phi(\theta)$ we have for an unbiased test

$$\left.\begin{aligned} \phi_\phi(\theta) &\leqslant \alpha, \quad \theta \in \omega \\ \phi_\phi(\theta) &\geqslant \alpha, \quad \theta \in \Omega \setminus \omega \end{aligned}\right\} \quad \ldots \quad (4.5)$$

(4.5) is a natural condition for the test ϕ to be considered as good. It does not guarantee by itself the test is good (for example, the trivial test $\phi \equiv \alpha$ is unbiased) but it seems quite reasonable.

Following E. Sverdrup [7] we remark that if the parameter space Ω is equipped with a topology for which $\phi_\phi(\theta)$ is a continuous function of θ and if the boundary $\partial\omega$ of the domains ω and $\Omega \setminus \omega$ is not empty then (4.5) implies that $\phi_\phi(\theta) = \alpha$ for $\theta \in \partial\omega$.

To fix ideas let us return to the exponential families considered in chapter one. Let $\{p_\theta(x)\}$ be an exponential family of the densities

$$p_\theta(x) = C(\theta)\, h(x) \exp(\theta_1 T_1 + \ldots + \theta_s T_s) \qquad \ldots \quad (4.6)$$

with respect to a σ-finite measure μ and let the parameter set be the natural one.

We shall be interested in testing the hypothesis

$$H_0 : \gamma_1(\theta_1, \ldots, \theta_s) = \gamma_{10}, \ldots, \gamma_q(\theta_1, \ldots, \theta_s) = \gamma_{q0},\ q < s, \qquad \ldots \quad (4.7)$$

$\gamma_1, \ldots, \gamma_q$ being sufficiently smooth functions. The alternative H_1 means the violation of at least one of the equalities (4.7).

The condition $q < s$ introduces the nuisance parameters which we have to bear in mind. We shall make the problem more precise assuming that $\gamma_1(\theta_1, \ldots, \theta_s)$ are entire functions and writing the relations (47) in the form

$$\left.\begin{array}{l} \pi_1(\theta_1, \ldots, \theta_s) = 0 \\ \ldots \quad \ldots \quad \ldots \\ \pi_q(\theta_1, \ldots, \theta_s) = 0 \end{array}\right\} \qquad \ldots \quad (4.8)$$

where the functions $\pi_1, \ldots, \pi_q$ are also entire. The hypothesis H_0 transforms the exponential family (4.6) with the natural parameters into an incomplete family with the relationships (4.8) which determine an analytic set (union of irreducible analytic varieties) ω in Ω. The alternative $H_1 : \theta \in \Omega \setminus \omega$ has as a boundary $\partial(\Omega \setminus \omega)$ in Ω a surface which is contiguous to ω.

If we have an unbiased test ϕ of the hypothesis H_0 then according to the argument by E. Sverdrup cited above ϕ must be similar because its power function is continuous in the standard topology of E_s. So we have

$$E_\theta \phi = \alpha \qquad \ldots \quad (4.9)$$

for θ satisfying the relations (4.8). If, besides $\pi_1, \ldots, \pi_q$, we can introduce the functions $\pi_{q+1}, \ldots, \pi_s$ such that $(\theta_1, \ldots, \theta_s) \to (\pi_1, \ldots, \pi_s)$ is a bijective homeomorphism then $\pi_{q+1}, \ldots, \pi_s$ will be nuisance parameters and the statistic $\phi = \phi(x)$ will be partially similar relative

to the parameters $\pi_{q+1}, \ldots, \pi_s$. If a partially similar test is in addition non-randomized then the statistic ϕ is (completely) similar relative to the nuisance parameters because the probability of the set where $\phi = 1$ (the critical region of the hypothesis H_0) will be independent of $\pi_{q+1}, \ldots, \pi_s$.

Hence we come again to the study of similar tests. If it is possible to describe all similar tests for a given problem of the type mentioned above then a test with the best quality can be chosen among them. Certainly the concept of "the best quality" should be defined first. Since the study of similar tests is rather difficult, we can first define the concept of quality in a simple manner by equipping the alternative H_1 with a Bayesian (*a priori*) measure $B(\theta)$. The optimal test relative to $B(\theta)$ is that for which

$$\int E_\theta \phi dB(\theta) \text{ is maximal.} \quad \ldots \quad (4.10)$$

If this test does not exist or if it is difficult to construct it we can try to find a similar "ϵ-optimal" test ϕ_ϵ for which

$$W(\phi_\epsilon/B) = \int_{\Omega \setminus \omega} E_\theta \phi_\epsilon \, dB(\theta) \geqslant \sup_{\Phi} \int_{\Omega \setminus \omega} E_\theta \phi d(B)(\theta) - \epsilon, \quad \ldots \quad (4.11)$$

where sup is taken over all similar tests of the given level α (see (4.9)).

Many problems with nuisance parameters specially in the multivariate analysis can be brought to the form considered above. It is worthwhile to mention the problem of testing the linear hypothesis when the variances are unknown (the case of least squares method with unknown weights, in particular the Behrens-Fisher problem described above).

REFERENCES

[1] E. L. Lehmann, *Testing Statistical Hypotheses*, N.Y., Wiley, 1959.

[2] G. Dantzig, On the non-existence of tests of Student's hypothesis having power function independent of σ, *Ann. Math. Stat.*, **XI**, 1940, pp. 186–191.

[3] Yu. V. Linnik, Statistical problems with nuisance parameters, Providence, *Amer. Math Soc.*, 1968.

[4] E. L. Lehmann, On the non-verifiability of certain parametric functions, *Teoria verojatn. i primen.*, **10**, 4 (1965), pp. 758–760.

[5] V. P. Palamodov, On verifiable functions (in Russian), *Teoria verojatn. i primen.*, **13**, 1 (1968).

[6] C. Stein, A two sample test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Stat.*, **16** (1945), pp. 259–281.

[7] E. Sverdrup, Similarity, unbiasedness, minimaxity and admissibiiity of statistical test procedures, *Skand. Aktuartidskriff*, **36**, 1953, pp. 64–86.

CHAPTER FIVE

COTEST : IDEALS OF COTESTS. THEOREM OF V. PALAMODOV

We return now to the exponential families (4.6) for which we shall study the problem of testing the hypothesis H_0 (4.8) with the aid of similar tests. The result due to Palamodov cited in chapter 4 shows that the verifiable functions are rather rare and one cannot expect that the functions (4.8) are verifiable : that is why we shall try to test the hypothesis (4.8) at least by similar tests.

For a given similar test $\phi(x)$ we can consider the conditional expectation

$$\phi(T) = E_\theta[\phi(x)\,|\,T], \qquad \ldots \quad (5.1)$$

$T = (T_1, \ldots, T_s)$ being a sufficient statistic for the family (4.6). $\phi(T)$ is independent of the parameters because of the sufficiency of T and $0 \leqslant \phi(T) \leqslant 1$. This means that $\phi(T)$ is again a test and in view of

$$E_\theta\phi(T) = E_\theta\phi(x) \qquad \ldots \quad (5.2)$$

$\phi(T)$ is a similar test if $\phi(x)$ was, i.e. if

$$E_\theta\phi(x) = \alpha, \quad \theta \in \omega$$

where ω denotes the analytic set (4.8). The relation (5.2) shows that $\phi(T)$ and $\phi(x)$ have the same power function and if we characterize the tests by their power functions we may restrict ourselves with the tests $\phi(T)$ depending only upon the sufficient statistic $T = (T_1, \ldots, T_s)$.

So from now on in studying the test $\phi(T)$ we shall be interested only in the sufficient statistic T and its distribution. It is well known [1] that the distributions of the vector $T = (T_1, \ldots, T_s) \in E_s$ form again an exponential family of densities

$$p(T, \theta) = C(\theta)h(T) \exp(\theta_1 T_1 + \ldots + \theta_s T_s) \qquad \ldots (5.3)$$

with respect to the Lebesgue measure on E_s.

The hypothesis H_0 introduces the relationships

$$\begin{aligned} &\pi_1(\theta_1, \ldots, \theta_s) = 0, \\ &\ldots \quad \ldots \quad \ldots \\ &\pi_q(\theta_1, \ldots, \theta_s) = 0, \end{aligned} \qquad \ldots \quad (5.4)$$

$\pi_1, \ldots, \pi_q$ being entire functions.

The relationships (5.4) define the set ω (of points of the hypothesis H_0).

For a similar test $\phi(T)$ with $E_\theta\phi = \alpha$ it is more convenient to deal with the function

$$\phi_c(T) = \phi(T) - \alpha \qquad \ldots \quad (5.5)$$

which we shall call cotest. For a function $\phi_c(T)$ to be a cotest it is necessary and sufficient that

$$-\alpha \leqslant \phi_c(T) \leqslant 1-\alpha \qquad \ldots \quad (5.6)$$

and

$$E_\theta\phi_c(T) = 0, \quad \theta \in \omega. \qquad \ldots \quad (5.7)$$

It is possible to describe all the cotests and then to construct with their aid the optimal or ϵ-optiaml tests in the sense of definitions (4.10), (4.11). With that end in view it is convenient to generalize a little the definition of cotest by omitting the condition (5.6) and preserving only the condition (5.7). We shall call a precotest any statistic $\xi(T)$ depending only on the sufficient statistic $T = (T_1, \ldots, T_s)$ such that $E_\theta\xi(T)$ exists and vanishes for $\theta \in \omega$. Hence for a precotest $\xi(T)$ we have

$$C(\theta) \int_{\mathcal{T}} \ldots \int \xi(T_1, \ldots, T_s) \exp(\theta_1 T_1 + \ldots + \theta_s T_s) h(T_1, \ldots, T_s) \times \\ \times dT_1, \ldots, dT_s = 0, \ \theta \in \omega \qquad \ldots \quad (5.8)$$

where $\mathcal{T} = \operatorname{supp} h$ denotes the support of h. The integral (5.8) is considered for the natural set of the parameterrs $\theta_1, \ldots, \theta_s$ where it converges absolutely. Here $C(\theta) \neq 0$ because if we set $\xi = 1$ the left-hand side of (5.8) will be equal to 1 for any point of the natural parameter set. To find all the precotests one can consider all statistics $m(T) = m(T_1, \ldots, T_s)$ for which the s-dimensional Laplace transform

$$\mathcal{L}(m \mid \theta) = \int_{E_s} \ldots \int m(T_1, \ldots, T_s) \exp(\theta_1 T_1 + \ldots + \theta_s T_s) dT_1, \ldots, dT_s \qquad \ldots \quad (5.9)$$

converges absolutely in a domain Θ of θ-values. We have to choose among them the statistics m with

$$\mathcal{L}(m \mid \theta) = 0, \quad \theta \in \omega. \qquad \ldots \quad (5.10)$$

If they vanish outside the set $\mathcal{T} = \text{supp } h$ then any function $\xi = \frac{m}{h}$ will be a precotest from which it is possible to construct a cotest if in addition there exists a constant A such that

$$-\alpha \leqslant A\xi(T) \leqslant 1-\alpha$$

for all values of T.

The statistics $m(T)$ for which the integral (5.9) converges absolutely for $\theta \in \Theta$ (where Θ contains a parallelopiped of positive measure) form a vector space $\mathcal{M}$. Let $m_1 \in \mathcal{M}$, $m_2 \in \mathcal{M}$; then the convolution

$$m_1 * m_2 = \int_{E_s} \ldots \int m_1(T_1-\xi_1, \ldots, T_s\xi_s) m_2(\xi_1, \ldots, \xi_s) d\xi_1, \ldots, d\xi_s \qquad \ldots \quad (5.11)$$

also belongs to $\mathcal{M}$ and we have

$$\mathcal{L}(m_1 * m_2 \mid \theta) = \mathcal{L}(m_1 \mid \theta)\mathcal{L}(m_2 \mid \theta).$$

The elements of $\mathcal{M}$ are thus equipped with a structure of the ring with the operation of convolution $*$ as multiplication. The images $\mathcal{L}(m \mid \theta)$ of the statistics $m \in \mathcal{M}$ also form a ring $\mathcal{L}(\mathcal{M})$ with the usual multiplication. The images of the statistics m statisfying (5.10) vanish on ω and hence they form an ideal of the ring $\mathcal{L}(\mathcal{M})$ which we shall call the cotest ideal of the hypothesis H_0 or simply the cotest ideal. It is only an abbreviation because the cotests themselves do not form an ideal but we shall use it. So to study the similar tests we have the following approach : we shall investigate the cotest ideals of the ring $\mathcal{L}(\mathcal{M})$ formed by the functions vanishing on the analytic set ω defined by the relationships (5.4). Having described these ideals we shall try then to obtain the inverse images of their elements by application of the inverse Laplace transform. We shall get in such a way the statistic $m(T)$ satisfying the condition (5.10). Among them we have to choose those which vanish outside supp h, i.e. those for which

$$\text{supp } m \subset \text{supp } h. \qquad \ldots \quad (5.12)$$

Then we construct the precotests $\xi = \frac{m}{h}$ and try to obtain from them the cotests. At last it will be possible to construct the optimal or ϵ-optimal similar tests in the sense of (4.10) and (4.11). This programme of determination of all the similar tests needs many rather complicated tools. For many problems of the kind under investigation the ring $\mathfrak{L}(\mathcal{M})$ is formed by the holomorphic functions in an appropriate domain (we suppose that the complexification of the parameter space is already done); in these situations the cotest ideals can be described with the aid of the theory of analytic sheaves developed by H. Cartan [2]*. One can also utilize the methods due to L. Hörmander [3] as will be shown later. The behaviour of the set supp h is of decisive importance. This way of description of the similar tests leads to rather simple formulae if we impose corresponding conditions.

This programme for rather general cases was realized in the monograph [4]. Though the final formulae are simple enough the methods and the necessary conditions are cumbersome. That is why we do not expound here the theory in the whole possible generality addressing the readers to [4], and consider only a special case which is general enough to solve completely the problem of description of similar tests. This special case includes among others the Behrens-Fisher problem for which all similar randomized tests will be described. We shall follow here V. Palamodov [5] expounding his method in detail.

We shall always consider the family (5.3) and the hypothesis (5.4). Before we formulate the conditions imposed on $h(T)$ and on the set Θ we introduce the set $\mathcal{X} \subset E_s$ defining it in the following way.

Consider the set of all those $\theta \in E_s$ for which there exists $c = c(\theta)$ such that

$$\{T : (\theta, T) < c\} \supset \mathcal{T} = \text{supp } h. \qquad \ldots \quad (5.13)$$

This set is a cone in E_s whose interior we denote by $\mathcal{X}$.

The following remark is important : for any $\theta \in \mathcal{X}$ one can choose a constant B such that for a suitable $\epsilon > 0$

$$(\theta, T) \leqslant -\epsilon\, |T| + B, \quad T \in J \qquad \ldots \quad (5.14)$$

* This method was chosen in the monograph [4].

where $|T|^2 = |T_1|^2 + \ldots + |T_s|^2$.

Let us formulate now the following conditions.

I. $\mathscr{T} = \operatorname{supp} h$ is a convex set.

II. $\mathscr{X}$ is not empty.

III. For any $\epsilon > 0$

$$\|h(T)\exp -\epsilon|T|\|_{L_2} < \infty, \qquad \ldots \quad (5.15)$$

where $\|\cdot\|_{L_2}$ denotes the L_2-norm.

Remark that (5.14) and (5.15) guarantee that for all $\theta \in \mathscr{X}$, $p(T, \theta)$ is a probability density.

IV. The parameter θ belongs to a set everywhere dense in $\mathscr{X} \bigcap \omega$ where ω is a real algebraic set.

V. The polynomial ideal I formed by all real polynomials of θ vanishing on $\mathscr{X} \bigcap \omega$ is a principal one, i.e. it consists of all polynomials divisible by a fixed polynomial $P(\theta)$.

Denote by N the variety of complex roots of $P(\theta)$ and set

$$\mathscr{X} \times E_s^{(y)} = \{\theta \in C^s, \quad \operatorname{Re} \theta \in \mathscr{X}\}, \qquad \ldots \quad (5.16)$$

C^s being the product of s complex planes. Note the relation

$$N \bigcap \mathscr{X} = \omega \bigcap \mathscr{X}.$$

VI. Each connected component of the intersection $N \bigcap (\mathscr{X} \times E_s^{(y)})$ has a least one real point where

$$\operatorname{grad} P(\theta) \neq 0.$$

Theorem 5.1. (V. Palamodov). *The conditions I–VI being satisfied, each cotest can be represented in the form*

$$\phi_c(T) = \frac{1}{h(T)} P(-D)\psi(T), \quad D = \left(\frac{\partial}{\partial T_1}, \ldots, \frac{\partial}{\partial T_s}\right), \quad \ldots \quad (5.17)$$

where $\psi(T)$ is an (ordinary) function such that

$$\left.\begin{aligned} &\operatorname{supp} \psi \subset \operatorname{supp} h \\ &\|\psi \exp -\epsilon|T|\|_{L_2} < \infty \text{ for any } \epsilon > 0. \end{aligned}\right\} \quad \ldots \quad (5.18)$$

The function ψ is uniquely determined.

Conversely each function of the form (5.17) with $\psi(T)$ satisfying the conditions (5.18) and lying between $-\alpha$ and $1-\alpha$ for a value of $\alpha \in (0, 1)$ is a cotest.

The proof of the theorem will be given in the next chapter.

REFERENCES

[1] E. L. Lehmann, *Testing Statistical Hypothesis*, N.Y., Wiley, 1959.

[2] H. Cartan, Variétés analytiques complexes et cohomologie (in French), Colloque sur les fonctions de plausieurs variables, Paris, Masson and Co., 1953.

[3] L. Hörmander, Linear partial differential operators, *Berlin, Akad. Press*, 1963.

[4] Yu. V. Linnik, Statistical problems with nuisance parameters, Providence, *Amer. Math. Soc.*, 1968.

[5] V. P. Palamodov, Testing multidimensional statistical hypotheses (in Russian), *Doklady Akad. Nauk SSSR*, **172**, 2 (1966).

CHAPTER SIX

THEOREM OF V. PALAMODOV (CONTINUATION)

We proceed now to the proof of Theorem 5.1.

Take an arbitrary point $a \in x$. Then for certain constants $c > 0$ and B the condition

$$(a, T) \leqslant -c|T|+B \qquad \ldots \quad (6.1)$$

is fulfilled for $T \in \operatorname{supp} h$.

Set

$$\|\phi\|_\varepsilon = \|\phi(T)\ \exp \epsilon|T|\|_{L_2}.$$

According to a theorem due to L. Hörmander [1] for any differential operator with constant coefficients there exists a fundamental solution $\epsilon(T)$* satisfying the condition

$$\|\epsilon * \phi\|_{-\varepsilon} \leqslant C\|\phi\|_\varepsilon \qquad \ldots \quad (6.2)$$

for any $\epsilon > 0$ and any ϕ for which the right-hand side of (6.2) is finite (the letters C, c with or without indices mean always positive constants). Recall that for the fundamental solution $\epsilon(T)$ the relation $D\epsilon(T) = \delta(T)$ holds, $\delta(T)$ being the Dirac delta-function.

Let $\epsilon(T)$ be a fundamental solution for the operator $P(-D+a)$. Consider a cotest ϕ_c with

$$E_\theta \phi_c = 0$$

for the values of θ indicated in the condition IV. Put,

$$\check{\phi} = \phi_c h$$

and construct the function

$$\begin{aligned}\psi(\tau) &= \epsilon(T) \exp[-(a, T)] * \check{\phi}(T) \\ &= \int \epsilon(T) \exp[-(a, T)]\, \check{\phi}(\tau - T)dT \\ &= \exp -(a, \tau) \int \epsilon(T) \exp(a, \tau - T)\, \check{\phi}(\tau - T)dT.\end{aligned}$$

* The solution $\epsilon(T)$ belongs to the space of generalized functions (see [1]).

In view of the inequality (6.1) the function $\exp(a, T)\,\check{\phi}(T)$ belongs to the space L_2 with the weight $\exp \epsilon |T|$ for a suitable $\epsilon > 0$. Then the property (6.2) of $\epsilon(T)$ implies that

$$\| \exp(a, \tau)\psi(\tau)\|_{-\varepsilon} = \|\epsilon * \exp(a, T)\check{\phi}(T)\|_{-\varepsilon}$$

$$\leqslant C \| \exp(a, T)\,\check{\phi}(T)\|_{\varepsilon} \leqslant C' \| \check{\phi} \|_{-\varepsilon'} \qquad \ldots \quad (6.3)$$

for all small $\epsilon' > 0$.

Let us verify the relation (5.17). To this end we shall establish first that the function $\epsilon(T) \exp -(a, T)$ is a fundamental solution for the operator $P(-D)$.

By the Leibnitz differentiation formula for the product we have

$$P(-D)\{\epsilon(T) \exp -(a, T)\} = \exp(-a, T) \sum_i \frac{a^i}{i!} P^{(i)}(-D)\epsilon$$

$$= \exp[-(a, T)]P(-D+a)\epsilon$$

$$= \exp[-(a, T)]\delta = \delta \qquad \ldots \quad (6.4)$$

from which easily follows

$$P(-D)\psi = P(-D)[\epsilon(T)\exp[-(a, T)] * \check{\phi}] = \delta * \check{\phi} = \check{\phi}.$$

We have to prove now that

$$\operatorname{supp} \psi \subset \operatorname{supp} h.$$

Let $\tau \bar{\in} \operatorname{Supp} h$. From the convexity of $\operatorname{supp} h$ it follows that there exists a vector λ such that

$$(\lambda, T) > (\lambda, \tau) \quad \text{for} \quad T \in \operatorname{supp} h.$$

The construction of $\epsilon(T)$ leads to the relation

$$P(-D+a)\,\epsilon(\tau - T) = 0 \text{ in the domain } (\lambda, \tau - T) < 0.$$

According to a theorem of V. Palamodov in the theory of differential operators with constant coefficients (see [1, 2]) for any $\epsilon > 0$ there exists a representation

$$E(\tau - T) = \int_{\substack{P(\theta+a)=0 \\ |Re\,\theta|<2\epsilon}} \exp(\theta, T)d\mu(\theta),$$

the measure μ depending on ϵ.

The measure μ is such that the integral in the right-hand side converges absolutely as a generalized function (distribution) in the half-space $(\lambda, \tau-T) < 0$.

We now have

$$\psi(\tau) = \int_{\mathcal{T}} E(\tau-T)\exp[-(a, \tau-T)]\check{\phi}(T)dT$$

$$= \exp(-a, \tau) \iint\limits_{\substack{P(\theta+a)=0 \\ |Re\,\theta|<2\epsilon}} \exp(\theta+a, T)\check{\phi}(T)dTd\mu(\theta). \quad \ldots \quad (6.5)$$

The interchanging of the order of integration is permissible because for a sufficiently small $\epsilon' > 0$ we have in view of $|\mathrm{Re}\,\theta| < 2\epsilon$

$$\exp(\theta+a, T) = 0(\exp-\epsilon'|T|), \quad T \in J.$$

We shall show now that

$$\int \exp(\theta+a, T)\check{\phi}(T)dT = 0$$

if

$$P(\theta+a) = 0, \quad |\mathrm{Re}\,\theta| < 2\epsilon.$$

For sufficiently small a we have $\theta+a \in \mathcal{X}\times E_s^{(y)}$. Let N' be a connected component of the intersection

$$N \bigcap (\mathcal{X}\times E_s^{(y)}),$$

Containing the point $\theta+a$. According to the condition VI, N' contains at least one real point z at which $\mathrm{grad}\,P(z) \neq 0$. Since $P(\theta)$ is a polynomial with real coefficients the condition $\mathrm{grad}\,P(z) \neq 0$ implies that N' contains an open $(S-1)$-dimensional subset ν of the real zeros of $P(\theta)$. The condition

$$E_0\phi = 0, \quad \theta \in \mathcal{X} \bigcap \omega$$

implies that

$$\int \exp(\theta, T)\check{\phi}(T)dT = 0, \quad \theta \in \nu. \quad \ldots \quad (6.6)$$

The function $\int \exp(\theta, T)\check{\phi}(T)dT$ is analytic in $\mathcal{X}\times E_s^{(y)}$. The principle of analytic continuation applied to (6.6) guarantees that the relation (6.6) holds on the whole set N', i.e.

$$\int \exp(\theta+a, T)\check{\phi}(T)dT = 0.$$

The relation (6.5) now implies $\psi(\tau) = 0$. Since τ was an arbitrary point outside of $\mathcal{T} = \operatorname{supp} h$ we see that

$$\operatorname{supp} \psi \subset \operatorname{supp} h.$$

Finally we show that for any $\epsilon > 0$

$$\| \psi \|_{-\epsilon} < \infty.$$

The relation (6.3) implies that

$$\| \psi \exp(a, T) \|_{-\epsilon} < \infty.$$

First we show that ψ does not depend on the choice of the point $a \in \mathcal{X}$. For two arbitary points $a, a' \in \mathcal{X}$ we set

$$\psi_a = \psi, \quad \psi_{a'} = \psi'.$$

We have

$$\| \psi \exp(a, T) \|_{-\epsilon} < \infty, \quad \| \psi' \exp(a', T) \|_{-\epsilon} < \infty. \qquad \ldots \quad (6.7)$$

From the convexity of the cone $\mathcal{X}$ and (6.7) follows the absolute convergence of the integrals

$$\tilde{\psi} = \int \psi \exp(\theta, T) dT \quad \text{and} \quad \tilde{\psi}' = \int \psi' \exp(\theta, T) dT.$$

for $\theta = b + a + a'$, $b \in \mathcal{X}$. Since

$$P(-D)(\psi - \psi') = 0$$

we have

$$P(\theta)[\tilde{\psi}(\theta) - \tilde{\psi}'(\theta)] = 0, \quad \theta \in \mathcal{X} + a + a',$$

$\mathcal{X} + a + a'$ being a cone moved by the vector $a + a'$. Then $\tilde{\psi} = \tilde{\psi}'$ and $\psi = \psi'$ from where

$$\| \psi \exp(a, T) \|_{-\epsilon} < \infty$$

for all $a \in \mathcal{X}$. Since we can choose $a \in \mathcal{X}$ with arbitrarily small $|a|$ we have

$$\| \psi \|_{-\epsilon} < \infty.$$

The arguments of this section show also that the function figuring in (5.17) and obeying the condition (5.18) is uniquely determined.

Let now $\phi_c(T)$ be a function given by the formula (5.17) with the conditions (5.18) and (5.19). Then we have for $\theta \in \mathcal{X}$

$$E_\theta \phi_c(T) = C(\theta) \int \cdots \int_\tau P(-D)\psi(T) \exp(\theta, T) dT$$

$$= C(\theta)P(\theta) \int \cdots \int_\tau \psi(T) \exp(\theta, T) dT \quad \ldots \quad (6.8)$$

according to the theory of Laplace transforms [4]. Since $P(\theta) = 0$ for $\theta \in \mathcal{X} \bigcap \omega$, we have $E_\theta \phi_c = 0$, $\theta \in \mathcal{X} \bigcap \omega$, i.e. ϕ_c is a cotest.

We conclude with a number of remarks.

The function ψ in (5.17) is an ordinary (not generalized) function allowing the application of the operator $P(-D)$ in the sense that the result of this application leads again to an ordinary function. But this does not mean that any part of the operator $P(-D)$ may be applied to ψ in the same sense as above. Hence we can not restrict ourselves in general with the functions $\psi(T)$ which are differentiable in the ordinary sense as many times as necessary for the application of $P(-D)$.

In case of elliptic operators the situation is simpler.

Another approach developed in the monograph [5] leads to the formulae in terms of convolutions, i.e. of integral operators instead of differential ones.

But the approach of [5] gives the description of all sufficiently smooth tests (it is enough if we are interested in ϵ-optimal tests, for example) though it needs rather cumbersome conditions.

We shall give an outline of the second approach. Suppose that the set $\mathcal{T} = \operatorname{supp} h$ coincides with the domain $\{T_i < 0,\ i = 1, \ldots, s\}$ and that for $\theta_i \to \infty$, $\operatorname{Re} \theta_i > 0$, we have

$$\int \cdots \int_\tau \psi(T) \exp(\theta, T) dT = 0\left(\frac{1}{(\theta_1, \ldots, \theta_s)^{k+2}}\right), \quad \ldots \quad (6.9)$$

the constant k being greater than the maximal degree m of the variables in $P(\theta)$. Then it is possible to write (6.8) in the form

$$E_\theta \phi(T) = \frac{C(\theta)\, P(\theta)}{(\theta_1, \ldots, \theta_s)^k} \int \cdots \int_\tau (\theta_1, \ldots, \theta_s)^k \psi(T) \exp(\theta, T) dT. \quad \ldots \quad (6.10)$$

The function $\dfrac{P(\theta)}{(\theta_1, \ldots, \theta_s)^k}$ is a Laplace transform of a polynomial $A(T)$ (for example, the image of $\left(-1\right)^k$ is $\dfrac{1}{\theta_1, \ldots, \theta_k}$). The function

$$\int \cdots \int_\tau (\theta_1, \ldots, \theta_s)^k \psi(T) \exp(\theta, T) dT$$

is a Laplace transform of a smooth function $F(T)$ with supp F in the domain $T_i < 0$, $i = 1, \ldots, k$. On the other hand $E_0\phi_c(T)$ is a Laplace transform of the function $\phi_c.h$. Thus we have

$$\phi_c = \frac{1}{h}(A * F) \qquad \ldots \quad (6.11)$$

which expresses the cotests in terms of the convolution operator. Here the condition on the set supp h is important. As we have remarked (cf. [5]) the condition (6.7) implies no loss in generality if we are interested only in ϵ-optimal similar tests. The condition on supp h, necessary for obtaining the formula (6.11), is cumbersome but it is satisfied in many problems, particularly, in multivariate analysis.

The general formula (5.17) or the formula (6.11) (when it is valid) can be useful in dealing with the optimization problems, for example, in the sense of (4.10) and (4.11). With the aid of (5.17), these optimization problems are reduced to certain variational problems, and the use of (6.11) allows us to reduce the optimization problems to certain problems of linear programming.

Palamodov's theorem can be generalized to the case of the exponential families with polynomial relationships of general type. In this case the ideal is not necessarily principal. Any cotest ϕ_c can be written in the form

$$\phi_c = \frac{1}{h} \sum_j P_j(-D)\psi_j$$

where $P_j(\theta)$ are generating elements of the ideal I and $\psi_j(T)$ are certain generalized functions with supports contained in $\mathcal{T}$.

The important condition of convexity of the set $\mathcal{T} = \text{supp } h$ is fulfilled at least in the problems of multivariate analysis where supp h is contained in the intersection of planes, ellipsoids and elliptic paraboloids of high dimensions. The Behrens-Fisher problem is one among them.

REFERENCES

[1] L. Hörmander, Linear partial differential operators, *Berlin, Akad. Press*, 1963.

[2] V. Palamodov, On systems of differential equations with constant coefficients (in Russian), *Doklady Akad. Nauk SSSSR*, **148**, 3 (1963).

[3] V. Palamodov, *Doctoral dissertation*, Moscow University, 1965.

[4] G. Doetsch, *Theorie und An. Wendung der Laplace-Transformation*, I, II, Basel, 1957.

[5] Yu. V. Linnik, Statistical problems with nuisance problems, Providence, *Amer. Math Soc.*, 1968.

CHAPTER SEVEN

APPLICATION OF THE V. PALAMODOV THEOREM TO THE BEHRENS-FISHER PROBLEM

The Behrens-Fisher problem mentioned in chapter 4 (example 4) is one of the most simple and well known ones in the theory of similar tests. Palamodov's theorem gives the complete soluton to the problem of description of all similar randomized tests for the Behrens-Fisher problem. As to the study of non-randomized tests for the Behrens-Fisher problem the theorem by V. Palamodov does not make it easier. We shall concern ourselves with the detailed study of non-randomized tests in lectures which follow where we shall construct similar tests for the Behrens-Fisher problem.

Let there be given two independent samples

$$x_1, \ldots, x_{n_1}, \quad x_i \sim N(a_1, \sigma_1^2)$$

$$y_1, \ldots, y_{n_2}, \quad y_j \sim N(a_2, \sigma_2^2)$$

and we want to test the hypothesis $a_1 = a_2$ with the help of similar (with respect to three nuisance parameters) tests.

There are four sufficient statistics

$$X_1 = \bar{x}, \quad X_2 = \bar{y}, \quad V_1 = s_1^2, \quad V_2 = s_2^2.$$

The well known procedures (see, for example [1] or [2]) lead to the following expression for the density of the random vector (X_1, X_2, V_1, V_2) at the point (x_1, x_2, v_1, v_2)

$$p(x_1, x_2, v_1, v_2) = C\sigma_1^{-n_1}\sigma_2^{-n_2} \exp\left(-\frac{n_1 a_1^2}{2\sigma_1^2} - \frac{n_2 a_2^2}{2\sigma_2^2}\right) \times$$

$$\times v_1^{\frac{n_1-3}{2}} v_2^{\frac{n_2-3}{2}} \exp\left[-\frac{n_1}{2\sigma_1^2}(v_1+x_1^2) - \frac{n_2}{2\sigma_2^2}(v_2+x_2^2) + \frac{n_1 a_1}{\sigma_1^2}x_1 + \frac{n_2 a_2}{\sigma_2^2}x_2\right] \quad \ldots \quad (7.1)$$

where $C > 0$.

Introduce the natural parameters :

$$\frac{n_1}{2\sigma_1^2} = \lambda_1, \frac{n_2}{2\sigma_2^2} = \lambda_2, \frac{n_1 a_1}{\sigma_1^2} = \mu_1, \frac{n_2 a_2}{\sigma_2^2} = \mu_2. \quad \ldots \quad (7.2)$$

We have then

$$2a_1 = \frac{\mu_1}{\lambda_1}, \quad 2a_2 = \frac{\mu_2}{\lambda_2}$$

so that the hypothesis $H_0 : a_1 = a_2$ can be written as

$$H_0 : \frac{\mu_1}{\lambda_1} - \frac{\mu_2}{\lambda_2} = 0. \quad \ldots \quad (7.3)$$

The more general hypothesis $H_\sigma : a_1 - a_2 = \delta$, δ being given, reduces to the hypothesis H_0 by passing from x_i to $x_i + \frac{\delta}{2}$, $i = 1, \ldots, n_1$ and from y_j to $y_j - \frac{\delta}{2}$, $j = 1, \ldots, n_2$. To transform the density (7.1) to the form (5.3) considered in the V. Palamodov's theorem we set

$$\frac{n_1 a_1}{\sigma_1^2} = \theta_1, \qquad -\frac{n_1}{2\sigma_1^2} = \theta_2,$$

$$\frac{n_2 a_2}{\sigma_2^2} = \theta_3, \qquad -\frac{n_2}{2\sigma_2^2} = \theta_4,$$

$$C\sigma_1^{-n_1}\sigma_2^{-n_2} \exp\left(-\frac{n_1 a_1^2}{2\sigma_1^2} - \frac{n_2 a_2^2}{2\sigma_2^2}\right) = C(\theta),$$

$$x_1 = T_1, \quad x_2 = T_3, \quad v_1 + x_1^2 = T_2, \quad v_2 + x_2^2 = T_4.$$

Then (7.1) can be rewritten as

$$p(T, \theta) = C(\theta)h(T) \exp(\theta_1 T_1 + \theta_2 T_2 + \theta_3 T_3 + \theta_4 T_4) \quad \ldots \quad (7.4)$$

where

$$h(T) = (T_2 - T_1^2)^{\frac{n_1 - 3}{2}} (T_4 - T_3^2)^{\frac{n_2 - 3}{2}}, \quad \ldots \quad (7.5)$$

$$T_1^2 \leqslant T_2, \quad T_3^2 \leqslant T_4, \quad \theta_2 < 0, \quad \theta_4 < 0.$$

The hypothesis H_0 (relation (7.3)) will take the form of

$$\theta_1\theta_4 - \theta_2\theta_3 = 0. \quad \ldots \quad (7.6)$$

The corresponding ideal I is a principal one and is generated by the polynomial

$$P(\theta) = \theta_1\theta_4 - \theta_2\theta_3.$$

One can easily verify conditions I–V; note in particular that supp h is convex. Let us verify now the condition VI. Take an arbitrary point

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4) \in N \times (\mathcal{X} \times E_s^{(\nu)})$$

and construct a continuous curve belonging to the set $N \bigcap (\mathcal{X} \times E_s^{(\nu)})$ which connects the point θ with the point (1, 1, 1, 1). With this end in view let us construct a continuous curve belonging to the product of the half-planes $\{\mathrm{Re}\ \theta_2 < 0\}$, $\{\mathrm{Re}\ \theta_4 < 0\}$ and the plane of θ_3-values which connects the points $(\theta_2, \theta_3, \theta_4)$ and (1, 1, 1). When the point $(\theta_2', \theta_3', \theta_4')$ runs over this curve the point $\left(\frac{\theta_2'\theta_3'}{\theta_4'}, \theta_2', \theta_3', \theta_4'\right)$ runs over the first curve on $N \bigcap (\mathcal{X} \times E_s^{(\nu)})$. It remains only to remark that for $P(\theta) = \theta_1\theta_4 - \theta_2\theta_3$ grad $P(\theta) = 0$ only at the origin which does not belong to $\mathcal{X}$.

Hence for the Behrens-Fisher problem all cotests are of the form

$$\phi_c(T) = \frac{1}{h(T)}\left(\frac{\partial^2}{\partial T_1 \partial T_4} - \frac{\partial^2}{\partial T_2 \partial T_3}\right)\psi(T), \qquad \ldots \quad (7.7)$$

$\psi(T)$ satisfying the condition (5.18).

As it is seen the operator $P(-D)$ is, here, a hyperbolic one, called a hyperwave operator.

It remains only to solve the problem of finding the optimal or ϵ-optimal (in sense of (4.9), (4.10), for example) tests with the aid of (7.7). It may happen that these tests are of a very complicated structure and they are inconvenient for application in practice. But in any case they may serve for comparing the quality of rather simple tests (as, for example, the Bartlett-Scheffé test (see [2]) with that of the optimal test.

REFERENCES

[1] E. L. Lehmann, *Testing Statistical Hypotheses*, N.Y., Wiley, 1959.

[2] Yu. V. Linnk, Statistical problems with nuisance parameters, Providence, *Amer. Math. Soc.*, 1968.

CHAPTER EIGHT

PROPERTIES OF HOMOGENEOUS SIMILAR TESTS FOR THE BEHRENS-FISHER PROBLEM

In chapter 7 we constructed all randomized similar tests for the Behrens-Fisher problem depending on sufficient statistics $\bar{x}, \bar{y}, s_1^2, s_2^2$. It is interesting to investigate among them the tests which are "homogeneous". We shall now make precise this notion.

If in the Behrens-Fisher situation we have $\sigma_1 = \sigma_2 = \sigma$ we come to the Student problem of testing the hypothesis $H_0 : a_1 = a_2$ on the base of two groups of independent observations : $x_1, \ldots, x_{n_1}$, $x_i \sim N(a_1, \sigma^2)$, and $y_1, \ldots, y_{n_2}$, $y_j \sim N(a_2, \sigma^2)$. The classical similar test of Student is given by the critical region

$$\left| \frac{\bar{x}-\bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{(n_1+n_2-2)(n_1+n_2)}{n_1 n_2}} \right| > C_0,$$

C_0 being a constant.

We see that the similar statistic of the test can be represented in the form

$$\frac{\bar{x}-\bar{y}}{s_2}\left[n_1\left(\frac{s_1}{s_2}\right)^2+1\right]^{-\frac{1}{2}},$$

i.e. it depends only upon the statistics $\frac{\bar{x}-\bar{y}}{s_2}$ and $\frac{s_1}{s_2}$. We shall call the tests of the form

$$\Phi\left(\frac{\bar{x}-\bar{y}}{s_2}, \frac{s_1}{s_2}\right)$$

homogeneous.

We shall show later that very simple axioms (A. Wald's axioms) lead in a natural manner to the tests of this form. If there is a

homogeneous test $\Phi\left(\frac{\bar{x}-\bar{y}}{s_2}, \frac{s_1}{s_2}\right)$ for the Behrens-Fisher problem it seems natural to impose the following condition : to accept the hypothesis $H_0 : a_1 = a_2$ if the value of $\frac{\bar{x}-\bar{y}}{s_2}$ is sufficiently small. In fact, to reject the hypothesis $H_0 : a_1 = a_2$ when the normed difference of sample means is small does not seem reasonable and the tests which imply the decisions of this kind are "irregular". Specifying this notion we introduce the following definition for the homogeneous tests $\Phi\left(\frac{\bar{x}-\bar{y}}{s_2}, \frac{s_1}{s_2}\right)$.

Property of null-regularity of a test $\Phi\left(\frac{\bar{x}-\bar{y}}{s_2}, \frac{s_1}{s_2}\right)$. Let $\xi = \frac{\bar{x}-\bar{y}}{s_2}$, $\eta = \frac{s_1}{s_2}$. There exists a number $\eta_0 = 1+\eta_1 > 1$ such that it is possible to draw a circle around every point (ξ, η) with $\xi = 0$, $0 \leqslant \eta < \eta_0$. such that when the observations place the point (ξ, η) in this circle the conditional probability of rejecting the hypothesis H_0 on the basis of test $\Phi(\xi, \eta)$ is zero.

If the test is non-randomized, i.e. if the function $\Phi(\xi, \eta)$ takes on only the values 0 and 1 and if the boundary of the ciritical region (where $\Phi = 1$) is connected and continuous then the null-regularity property of $\Phi(\xi, \eta)$ means simply that the test $\Phi(\xi, \eta)$ does not reject the hypothesis H_0 when $\bar{x}-\bar{y} = 0$ and $s_1^2+s_2^2 > 0$, which is an extremely simple and natural condition.

We are able to formulate now the following theorem (see [1, 2]) in which the condition $\max(n_1, n_2) \geqslant 3$ is assumed to be true.

Theorem 8.1. *There are no randomized homogeneous similar (and a fortiori unbiased) tests for the Behrens-Fisher problem which possess the null-regularity property.*

This theorem shows that the very natural property of a homogeneous test of being null-regular is incompatible with its unbiasedness (and even with its similarity) which also seems a natural and reasonable property. So the theorem 8.1 is of negative character.

It is natural to pose the question about the existence of homogeneous similar tests. The formula (7.7) due to V. Palamodov does not

allow to answer the question immediately. But the examples of such tests can be constructed in a rather simple manner. It might be noted (and we shall see later) that the probability density of the statistics (ξ, η) is an even function of ξ for all η if the hypothesis H_0 is true. Let $f_1(\xi, \eta)$ denote a continuous function of ξ and η in the half-plane Ω :

$$-\infty < \xi < \infty, \quad 0 \leqslant \eta < \infty$$

which is an odd function of ξ for each value of η and satisfies the condition

$$|f_1(\xi, \eta)| \leqslant \frac{1}{3}.$$

Let us set

$$f(\xi, \eta) = \tfrac{1}{2} + f_1(\xi, \eta).$$

Evidently $f(\xi, \eta)$ is a test and we have

$$Ef_1(\xi, \eta) = 0, \quad Ef(\xi, \eta) = \tfrac{1}{2}$$

if H_0 is true. Hence $f(\xi, \eta)$ is a similar randomized test. The null-regularity property is violated here : for $\xi = 0$, $f(\xi, \eta) = \frac{1}{2}$ and we reject the hypothesis H_0 with probability $\frac{1}{2}$. Other examples of certain practical value are given in [2, ch. 9]. In these examples the properties of independent statistics stated in chapter 2 find their application.

The detailed proof of the theorem 8.1 is given in [1, 2]; it requires many analytical transformations and calculations. We expound here only the idea of the proof and the basic relations referring the readers to [1, 2] for the details.

We begin with obtaining the distribution of the random vector

$$(\xi, \eta) = \left(\frac{\bar{x} - \bar{y}}{s_2}, \frac{s_1}{s_2} \right)$$

assuming that $n_1 \geqslant 2$, $n_2 \geqslant 2$. Set

$$X_1 = (\bar{x} - \bar{y}) \left(\frac{n_1 \sigma_2^2 + n_2 \sigma_1^2}{n_1 n_2} \right)^{-\frac{1}{2}}.$$

This variable is independent of s_1^2 and s_2^2 (see chapter 2). We have $EX_1 = 0$; $\mathfrak{D}X_1 = 1$, i.e. $X_1 \sim N(0, 1)$. Consider also the variables $u = \frac{n_1 s_1^2}{\sigma_1^2}$ and $v = \frac{n_2 s_2^2}{\sigma_2^2}$ which are independent and distributed according to the chi-square distribution with (n_1-1) and (n_2-1) degrees of freedom respectively. Let $u_1 = u^{\frac{1}{2}}$, $v_1 = v^{\frac{1}{2}}$. The variables X_1, u_1, v_1 are independent with densities

$$p(X_1) = \frac{1}{\sqrt{2\pi}} \exp - X_1^2/2,$$

$$q(u_1) = \frac{1}{2^{\frac{n_1-3}{2}} \Gamma\left(\frac{n_1-1}{2}\right)} u_1^{n_1-2} \exp -u_1^2/2, \; u_1 > 0,$$

$$r(v_1) = \frac{1}{2^{\frac{n_1-3}{2}} \Gamma\left(\frac{n_2-1}{2}\right)} v_1^{n_2-2} \exp -v_1^2/2, \; v_1 > 0.$$

Introduce the new variables

$$\tilde{x} = \frac{X_1}{v_1}, \quad \tilde{y} = \frac{u_1}{v_1}, \quad \tilde{z} = v_1.$$

Then

$$X_1 = \tilde{x}\tilde{z}, \quad u_1 = \tilde{y}\,\tilde{z}, \quad v_1 = \tilde{z},$$

$$\frac{\partial(X_1, u_1, v_1)}{\partial(\tilde{x}, \tilde{y}, \tilde{z})} = \tilde{z}^2$$

and elementary calculations show that the density of $(\tilde{x}, \tilde{y})$ is

$$\rho(\tilde{x}, \tilde{y}) = \tilde{C}_{n_1, n_2} \frac{\tilde{y}^{n_1-2}}{(\tilde{x}^2+\tilde{y}^2+1)^{(n_1+n_2-1)/2}}$$

where

$$\tilde{C}_{n_1, n_2} = \frac{2\Gamma\left(\frac{n_1+n_2-1}{2}\right)}{\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)}$$

Setting $\theta = \frac{n_2\sigma_1^2}{n_1\sigma_2^2}$ we shall have

$$\tilde{x} = \frac{\xi}{\sqrt{1+\theta}}, \quad \tilde{y} = \frac{\eta}{\sqrt{\theta}}.$$

The power function of a test $\Phi(\xi, \eta)$ can be now written as

$$E_\theta \Phi = \iint_\Omega \phi(\tilde{x}\sqrt{1+\theta}, \tilde{y}\sqrt{\theta})\, \rho(\tilde{x}, \tilde{y}) d\tilde{x}\, d\tilde{y}. \quad \ldots (8.1)$$

Here Ω denotes the upper half-plane : $-\infty < \tilde{x} < \infty$, $\tilde{y} \geqslant 0$. The condition of similarity is

$$E_\theta \phi = \alpha, \quad \theta \in H_0, \quad \ldots (8.2)$$

where α is a fixed level of significance.

Setting $N = \frac{n_1+n_2-1}{2}$ and reverting to the variables ξ, η, we rewrite the conditions (8.2) as

$$\iint_\Omega \Phi(\xi, \eta) \frac{\eta^{n_1-2} d\xi d\eta}{[\theta^2+\theta(1+\xi^2+\eta^2)+\eta^2]} = C(n_1, n_2)\alpha\theta^{-\frac{n_2}{2}}(1+\theta)^{-N+\frac{1}{2}} \quad \ldots (8.3)$$

where $C(n_1, n_2) > 0$.

The relation (8.3) is a basic integral relation of the theory of homogeneous similar tests for the Behrens-Fisher problem.

The main idea of investigating the relation (8.3) presented in this and next chapter is the making use of the analytic continuation of (8.3) for all possible values of θ in the complex plane. The denominator

$$\theta^2+\theta(1+\xi^2+\eta^2)+\eta^2 \quad \ldots (8.4)$$

in the left-hand side of (8.3) may vanish for the negative values of $\theta = -\aleph \leqslant 0$ along certain second order curves. We shall call these curves the critical curves of the family of distributions figuring in (8.3). They play an essential role in the study of homogeneous similar tests which we shall try to demonstrate.

For $\aleph = 0$ the critical curve is $\eta = 0$ (which we take for the abscissa axis).

For $0 < \aleph < 1$ we obtain the family of confocal hyperbolae

$$\frac{\eta^2}{\aleph} - \frac{\xi^2}{1-\aleph} = 1 \qquad \ldots (8.5)$$

with common focus $(0, 1)$; the axis $\eta = 0$ may be considered also as a limiting hyperbola.

For $\aleph = 1$ we have $\xi = 0$, i.e. the ordinate axis. The part of this axis, namely the twice covered ray $\xi = 0$, $1 \leqslant \eta < \infty$ also may be considered as a hyporbola of our family.

Hence we obtained the family of confocal hyperbolas

$$A(\xi, \eta) = \aleph, \quad 0 \leqslant \aleph \leqslant 1 \qquad \ldots (8.6)$$

which degenerate into segments for $\aleph = 0$, $\aleph = 1$.

For $\aleph > 1$ we get the family of ellipses

$$\frac{\xi^2}{\aleph - 1} + \frac{\eta^2}{\aleph} = 1 \qquad \ldots (8.7)$$

with common focus $(0, 1)$. When $\aleph = 1$ they degenerate into the twice covered segment $\xi = 0$, $0 \leqslant \eta \leqslant 1$.

The family of ellipses can be written in the form

$$B(\xi, \eta) = \aleph, \quad \aleph \geqslant 1. \qquad \ldots (8.8)$$

We pass now to the proof of the theorem 8.1 starting from the relation (8.3). Suppose that $\Phi(\xi, \eta)$ is a similar null-regular test of level α. By making use of (8.6) and (8.8) we get the decomposition.

$$\theta^2 + \theta(1 + \xi^2 + \eta^2) + \eta^2 = [\theta + A(\xi, \eta)]\,[\theta + B(\xi, \eta)]. \qquad \ldots (8.9)$$

Let us now carry out the analytic continuation of (8.3) with respect to $\theta = \sigma + i\tau$ into the domain Λ, which is the complex plane with a cut along the negative axis $\tau = 0$, $-\infty < \sigma < 0$. Since the both sides of (8.3) are holomorphic functions in Λ and (8.3) holds for positive values of θ, the relation (8.3) will remain true in Λ. Note here that the same conclusion would follow from the weaker assumption that $\Phi(\xi, \eta)$ is similar for a countable boundary set of the values of $\frac{\sigma_1}{\sigma_2}$ (and hence of the values of θ).

We can write the relation (8.3) for $\theta \in \Lambda$ in the following form

$$\iint_{\Omega} \Phi(\xi, \eta) \frac{\eta^{n_1-2} d\xi d\eta}{[(\theta+A(\xi,\eta))(\theta+B(\xi,\eta))]^N} = \rho'_{n_1, n_2} \alpha \theta^{-\frac{n_2}{2}} (1+\theta)^{-N+\frac{1}{2}} \quad \ldots (8.10)$$

The analytic continuation for $\theta \in \Lambda$ must take into account the possible ramification points.

If the test $\mathfrak{A}(\xi, \eta)$ possesses the null-regularity property then there exists a number $\eta_0 = 1+\eta_1$, such that it is possible to draw a circle around every point of the segment $\xi = 0,\ 0 \leqslant \eta \leqslant \eta_0$ in which $\Phi(\xi, \eta) = 0$ almost everywhere with respect to Lebesgue measure.

By the Heine-Borel theorem we can choose a finite covering of the segment $\xi = 0,\ 0 \leqslant \eta \leqslant \eta_0$ consisting of some of these circles. Thus we obtain an open domain Γ containing our segment such that $\phi(\xi, \eta) = 0$ almost everywhere in Γ. Consider now the family of semi-ellipses which are obtained from (8.7) when $\eta \geqslant 0$. If the number $\mathfrak{A} < 1$ is sufficiently close to 1 the corresponding semi-ellipse $B(\xi, \eta) = \mathfrak{A}$ will lie entirely in the domain Γ. Let us enlarge it now in the following manner. Let $\mathfrak{A}_0 < 1$ denote the least upper bound of those $\mathfrak{A}$ for which $\phi(\xi, \eta) = 0$ almost everywhere in the interior of the semi-ellipse (8.7). If $\mathfrak{A}_0 = \infty$ the test $\Phi(\xi, \eta)$ is trivial; so we have to consider only the case of $\mathfrak{A}_0 < \infty$. Let now z_0 be a small positive number (then $z_1, z_2, \ldots$ will be so as well). Set

$$\mathfrak{A}_0 = 1+z_0, \quad z_1 = \min\left(\frac{z_0}{2}, \frac{1}{2}\right)$$

and let us take the values of θ in the rectangle $Q \subset \Lambda$:

$$\theta = D_0+\delta+iz, \quad 0 \leqslant \delta \leqslant z_1, \quad 0 < z \leqslant z_1. \quad \ldots (8.11)$$

Consider now the relation (8.10) having written it in the form

$$\iint_{\Omega \setminus \Gamma} \Phi(\xi, \eta) \frac{\eta^{n_1-2}\, d\xi d\eta}{[(\theta+A(\xi,\eta))(\theta+B(\xi,\eta))]^N} = C'_{n_2, n_1} \alpha \theta^{-\frac{n_2}{2}} (1+\theta)^{-N+\frac{1}{2}},$$

$$\ldots (8.12)$$

$0 \leqslant A(\xi, \eta) \leqslant 1$. For $\theta \in Q$ we obtain

$$|\theta+A(\xi, \eta)| \geqslant D_0-z_1-1 \geqslant z_1 \quad \ldots (8.13)$$

by virtue of the definition of z_1. If the number z_0 (and consequently z_1) is small enough, we shall have for $\theta \in Q$ that the denominator in the left-hand side of (8.12) becomes small. This implies that the right-hand side of (8.12) should be large enough whereas it is bounded for $\theta \in Q$. This effect will be intensified if we apply to both sides of (8.12) the operator $L_0^{(k)}$ of differentiating k times on θ. If k is sufficiently large we obtain the contradiction; this is the main idea of the proof of the theorem 8.1. The details (rather cumbersome) of the proof are given in [1, 2].

REFERENCES

[1] Yu. V. Linnik, Randomized homogeneous tests for the Behrens-Fisher problem (in Russian), *Izvestija Akad. Nauk SSSR*, **28**, 2 (1964).

[2] Yu. V. Linnik, Statistical problems with nuisance parameters, Providence, *Amer. Math Soc.*, 1968.

CHAPTER NINE

ON THE FISHER-WELCH-WALD TEST

We pass now to the study of non-randomized homogeneous similar tests for the Behrens-Fisher problem. In his posthumous article [1] A. Wald gave a system of axioms which are very natural from the statistical point of view and which single out some of these tests.

The four axioms due to A. Wald, which follow, determine the critical region Z of a homogeneous test. The condition of similarity will be an additional axiom.

I. The critical region Z lies in the space of sufficient statistics $\bar{x}, \bar{y}, s_1^2, s_2^2$, i.e. if a sample $(x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2})$ belongs to Z then any other sample $(x_1', \ldots, x_{n_1}'; y_1', \ldots, y_{n_2}')$ with $\bar{x}' = \bar{x}, \bar{y}' = \bar{y}, s_1'^2 = s_1^2$ $s_2'^2 = s_2^2$ also belongs to Z.

II. If a sample $(x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \in Z$ then for any constant c

$$(x_1+c, \ldots, x_{n_1}+c;\ y_1+c, \ldots, y_{n_2}+c) \in Z.$$

III. If $(x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \in Z$ then for any $k \neq 0$

$$(kx_1, \ldots, kx_{n_1}; ky_1, \ldots, ky_{n_2}) \in Z.$$

IV. If $(x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \in Z$ and $(x_1', \ldots, x_{n_1}';\ y_1', \ldots, y_{n_2}')$ is another sample with

$$|\bar{y}'-\bar{x}'| > |\bar{y}-\bar{x}|,\ s_1'^2 = s_1^2,\ s_2'^2 = s_2^2,$$

then

$$(x_1', \ldots, x_{n_1}';\ y_1', \ldots, y_{n_2}') \in Z.$$

These four axioms imply that the critical region of a test must be of the form

$$\frac{|\bar{x}-\bar{y}|}{s_2} \geqslant \phi\left(\frac{s_1}{s_2}\right) \qquad \ldots \quad (9.1)$$

where ϕ is a single-valued measurable function.

In fact axiom I guarantees that the region Z is determined only by the values of $\bar{x}, \bar{y}, s_1^2, s_2^2$. It follows from axiom II that if $(\bar{x}, \bar{y}, s_1^2, s_2^2) \in Z$ then $(\bar{x}, -\bar{y}, 0, s_1^2, s_2^2) \in Z$ i.e. Z is determined only by the variables $\bar{x}-\bar{y}$, s_1^2, s_2^2. Setting $k = -1$ we get from axiom III that Z is symmetric, being determined by the values of $|\bar{x}-\bar{y}|$, s_1^2, s_2^2.

Let now $k = \frac{1}{s_2}$; we see that for determining Z all we need are the variables $\xi = \frac{|\bar{x}-\bar{y}|}{s_2}$ and $\eta = \frac{s_1}{s_2}$ and the corresponding test is homogeneous. One can represent Z in the plane of variables $\xi = \frac{\bar{x}-\bar{y}}{s_2}$, $\eta = \frac{s_1}{s_2}$; because of the symmetry of Z it is possible to consider Z only in the quadrant $\Omega_1 : \xi \geqslant 0, \eta \geqslant 0$.

Let us show now that every straight line $\eta = \eta_0$ ($\eta_0 > 0$) intersects the boundary of Z in the quadrant Ω_1 at no more than one point. If it is not the case there exist two points $(\xi_1, \eta_0) \in Z$ and $(\xi_2, \eta_0) \in Z$ with $\xi_2 > \xi_1 \geqslant 0$. Then there exists a point (ξ_3, η_0) with $\xi_1 < \xi_3 < \xi_2$ such that $(\xi_3, \eta_0) \in Z$. Let $(\bar{x}, \bar{y}, s_1^2, s_2^2)$ and $(\bar{x}', \bar{y}', s_1'^2, s_2'^2)$ be the values of sufficient statistics corresponding to the points (ξ_1, η_0) and (ξ_2, η_0). We have

$$\frac{s_1}{s_2} = \frac{s_1'}{s_2'}, \quad \frac{\bar{x}'-\bar{y}'}{s_2'} > \frac{\bar{x}-\bar{y}}{s_2}.$$

Let $k = \frac{s_1}{s_2}$ and take a point with the values of sufficient statistics $k\bar{x}, k\bar{y}, ks_1, ks_2$. It belongs to Z. Furthermore we have

$$ks_2 = s_2', \quad \frac{\bar{x}'-\bar{y}'}{s_2'} > \frac{k\bar{x}-k\bar{y}}{ks_2}.$$

According to axiom IV the point (ξ_3, η_0) must also belong to Z which contradicts the assumption made above. Thus the critical region Z is in Ω_1 of the form

$$\frac{|\bar{x}-\bar{y}|}{s_2} \geqslant \phi_1\left(\frac{s_1}{s_2}\right) \qquad \ldots \quad (9.2)$$

where ϕ is a single valued measurable function with possible infinite values. If we exclude this possibility and make use of only the functions ϕ in (9.2) assuming only finite values we obtain the Fisher-Welch-Wald tests called in honour of the authors who investigated them. These tests may be written in the form

$$\frac{|\bar{x}-\bar{y}|}{\sqrt{s_1^2+s_2^2}} \geqslant \phi_1\left(\frac{s_1}{s_2}\right) \qquad \ldots \quad (9.3)$$

which we deduce from (9.2) by setting

$$\phi_1 = \phi\left(1+\frac{s_1^2}{s_2^2}\right)^{\frac{1}{2}}.$$

The bibliography references regarding these tests can be found in the survey by Breny [2].

In his article [1] A. Wald constructed approximately similar tests of the form (9.2) where ϕ is a rational function and raised the question of constructing an exactly similar test for which the function ϕ is analytic for $\eta = \frac{s_1}{s_2} > 0$. We shall see later that such a test is impossible and even a stronger theorem can be formulated.

Consider the quadrant $\Omega_1 : \xi \geqslant 0, \eta \geqslant 0$ and following the notations of chapter 8 write the test (9.2) in the form

$$\xi \geqslant \phi(\eta). \qquad \ldots \quad (9.4)$$

We are able to formulate now the following theorem.

Theorem 9.1. *Suppose that $\phi(\eta)$ is continuous for $0 \leqslant \eta \leqslant 1$ and its derivative is finite for $0 \leqslant \eta \leqslant 1$, where $\phi'(0)$ means the right-hand derivative. Suppose furthermore that $\phi(\eta)$ satisfies the Lipschitz condition with exponent 1 for $1 \leqslant \eta \leqslant 2(M+1)$ where $M = \sup_{0 \leqslant \eta \leqslant 1} \phi(\eta)$. Then the test cannot be similar if $n_2 \geqslant 4$.*

We see that if $\xi = \phi(\eta)$ is the boundary of a similar test, $\phi(\eta)$ may be neither analytic nor differentiable in the semi-open interval $[0, 1)$, satisfying at the same time the Lipschitz condition in the finite interval figuring in the theorem.

The proof of the theorem is based on the principle of analytic continuation applied in the same manner as in chapter eight. The proof

includes many transformations and in particular very difficult analytic techniques. All the details of the proof are given in [3, ch. 8]. Here we shall demonstrate only the main ideas omitting the purely analytic calculations. Let $\alpha \in (0, 1)$ denote the level of a similar test of the form (9.4). First of all the following lemma can be established.

Lemma 1. Let $\xi_0 = \psi(0)$. Then we have

$$c_2\alpha \leqslant \xi_0 \leqslant c_1\alpha \qquad \ldots \quad (9.5)$$

($c_1, c_2, \ldots$ are positive constants).

The role of sample sizes n_1 and n_2 is not symmetric because the form of tests is not. But it is always possible to change the two samples by passing from one test to the other with the aid of the operation of "upsetting" which we describe now. For a given test

$$|\xi| \geqslant \phi(\eta) \qquad \ldots \quad (9.6)$$

let

$$\phi_1(\eta) = \frac{\phi(\eta)}{\sqrt{1+\eta^2}}.$$

Then $\phi(\eta) = \phi_1(\eta)\sqrt{1+\eta^2}$ and we may rewrite (9.6) as

$$|\xi| \geqslant \phi_1(\eta)\sqrt{1+\eta^2}$$

or

$$\frac{|\bar{x}-\bar{y}|}{s_2} \geqslant \phi_1\left(\frac{s_1}{s_2}\right)\sqrt{1+\frac{s_1^2}{s_2^2}}.$$

which is equivalent to

$$\frac{|\bar{y}-\bar{x}|}{s_1} \geqslant \phi_1\left(\left(\frac{s_2}{s_1}\right)^{-1}\right)\sqrt{1+\frac{s_2^2}{s_1^2}}$$

This last expression gives us the critical region of the test in which the roles of the sample $(x_1, \ldots, x_{n_1})$ is played by $(y_1, \ldots, y_{n_2})$ and *vice versa*. The test $|\xi| \geqslant \phi_1(\eta^{-1})\sqrt{1+\eta^2}$ which corresponds to the test $|\xi| \geqslant \phi_1(\eta)\sqrt{1+\eta^2}$ is of the same level α. This is the original operation of "upsetting" the tests.

It is interesting to give a geometric interpretation of this operation and to observe the behaviour of the critical curves introduced in chapter 8.

Consider the projective transformation of the plane

$$-\infty < \xi < \infty, \quad -\infty < \eta < \infty$$

defined by the formula $\xi' = \frac{\xi}{\eta}$, $\eta' = \frac{1}{\eta}$. This is an involution of the plane in the sense of projective geometry with the axis $\eta = 1$ and the centre $(0, -1)$. The quadrants of the upper half plane are transformed into themselves. The boundary $\xi = \phi_1(\eta)\sqrt{1+\eta^2}$ of the critical region in the upper right quadrant is transformed into $\xi = \phi_1(\eta^{-1})\sqrt{1+\eta^2}$ which corresponds to a test of the same level. The "upsetting" is an involutive operation.

Let $\alpha \in (0, 1)$ and $A(\xi, \eta) = \alpha$ be a critical hyperbola

$$\frac{\eta^2}{\alpha} - \frac{\xi^2}{1-\alpha} = 1.$$

The involution described above transforms the hyperbola into the ellipse

$$\frac{\xi^2}{\alpha^{-1}-1} + \frac{\eta^2}{\alpha^{-1}} = 1$$

which is also a critical curve.

The following lemma plays a fundamental role for the proof of the theorem 9.1. It is a lemma on the tangency point.

Lemma 2. Let (ξ_0, η_0) be a tangency point between the boundary $\xi = \phi(\eta)$ of a critical region and a nondegenerate critical curve. Suppose that $\phi(\eta)$ has a finite derivative $\phi'(\eta_0)$ and $(0, \eta'_0)$ denotes the point of intersection the vertical axis $\xi = 0$ by the critical curve tangent to $\xi = \phi(\eta)$. If $\phi(\eta)$ satisfies the Lipschitz condition in a neighbourhood of the point η'_0 then the test $|\xi| \geqslant \phi(\eta)$ cannot be a similar one.

We indicate the main ideas of the proof of Lemma 2. We start from the integral relation (8.3) which may be written in the form

$$\iint_{Z_1} \frac{\eta^{n_1-2}d\eta}{[(\theta+A(\xi,\eta))]\,[(\theta+B(\xi,\eta))]^N} = C_{n_1,n_2}\frac{\alpha}{2}\theta^{-\frac{n_2}{2}}(1+\theta)^{-N+\frac{1}{2}} \quad \ldots \quad (9.7)$$

if one takes into account the decomposition of the denominator as in Chapter 8 and non-randomizedness of the test. Z_1 denotes here the part of the critical region placed in the first quadrant.

It may be assumed that the tangency takes place with a critical hyperbola $A(\xi, \eta) = \aleph_0$, $0 < \aleph_0 < 1$. If it is not the case then it is possible to "upset" the original test and the new test will be tangent to a critical hyperbola.

Consider now the values of θ of the form $\theta = -\aleph_0 + iz$ where z is a small real number. Then for θ belonging to the complex plane Λ with a cut introduced in the previous chapter the denominator in the left-hand side integral of (9.7) becomes small whereas the right-hand side of (9.7) is bounded in a neighbourhood of $\theta = -\aleph_0$. To obtain the contradiction not at the heuristic but at a strong mathematical level in [3, ch. 8] a rather complicated integro-differential operator was introduced which when applied to the both parts of (9.7) leads to the contradiction for a period $\theta = -\aleph_0 + iz$. We omit the difficult calculations connected with the application of the method.

The following lemma is much more simple.

Lemma 3. If the boundary $\xi = \phi(\eta)$ of a critical region is continuous on $[0, \eta_1]$, $\eta_1 > 1$ the function $\phi(\eta)$ has a zero in $(0, 1]$.

In Lemma 1 we have seen that $\phi(0) = \xi_0 \neq 0$. If the function $\phi(\eta)$ does not vanish in $(0, 1]$ it will not vanish in $(0, \eta_1]$ for a certain $\eta_1 > 1$ because of the continuity of $\phi(\eta)$ in a neighbourhood of $\eta = 1$.

Thus the test is null-regular and it may not be similar according to Theorem 8.1.

Lemma 4. Let $\phi(\eta)$ be a continuous function in $[0, 1)$ with a finite derivative $\phi'(\eta)$ for $0 \leqslant \eta < 1$. If $n_2 \geqslant 4$ then the curve $\xi = \phi(\eta)$ is tangent to a non-degenerate critical ellipse at a point (ξ_0, η_0), where $0 < \eta_0 < 1$.

From Lemma 1 we have $\phi(0) = \xi_0 > 0$. Suppose that

$$\phi'(0) > 0. \qquad \ldots \quad (9.8)$$

Then the curve $\xi = \phi(\eta)$ starting from the point $(\xi_0, 0)$ has a right hand tangent. Since $\phi(\eta)$ vanishes somewhere in the interval $(0,1]$ the curve $\xi = \phi(\eta)$ must turn to the left to revert to a point $(0, \eta'_0)$, $0 < \eta'_0 \leqslant 1$ at which $\phi(\eta)$ vanishes. It is clear that the curve must, be tangent to a non-degenerate critical ellipse $B(\xi, \eta) = \mathfrak{A}$, $\mathfrak{A} > 1$ that contains the loop of the curve $\xi = \phi(\eta)$ between $\eta = 0$ and $\eta = \eta'_0$. If $M = \sup_{0 \leqslant \eta \leqslant 1} \phi(\eta)$ we can find an upper bound for the ordinate η_2 of the intersection point of the ellipse $B(\xi, \eta) = \mathfrak{A}$ with the axis $\xi = 0$.

Indeed let $\frac{\xi^2}{\mathfrak{A}-1} + \frac{\eta^2}{\mathfrak{A}} = 1$ be the equation of the critical ellipse. Then one can easily see that

$$\eta_2 \leqslant \frac{3}{2}(M+1).$$

It remains to prove the inequality (9.8). The proof consists of a statistical argument and of many purely analytic calculations (see [3, ch. 8]). We shall demonstrate here only the statistical part of the proof. The acceptance region of the hypothesis $H_0 : a_1 = a_2$ is $|\xi| \leqslant \phi(\eta)$ or

$$\frac{|\bar{x}-\bar{y}|}{s_2} \leqslant \phi\left(\frac{s_1^2}{s_2^2}\right). \qquad \ldots \quad (9.9)$$

Introducing the same variables as in Chapter 8 we may rewrite (9.9) as

$$|X_1| \leqslant \frac{\sqrt{\chi'^2_{n_2-1}}}{\sqrt{1+\theta}} \phi\left(\theta \frac{\chi^2_{n_2-1}}{\chi'^2_{n_2-1}}\right) \qquad \ldots \quad (9.10)$$

where $X_1 \sim N(0, 1)$, $\chi^2_{n_1-1}$ and $\chi'^2_{n_2-1}$ are independent variables following chi-square distributions with n_1-1 and n_2-1 degrees of freedom respectively and $\theta = \frac{n_2\sigma_1^2}{n_1\sigma_2^2}$.

Set

$$G(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt = P(|X_1| \leqslant 4),$$

$$g(u) = \frac{2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

If $\alpha \in (0, 1)$ is a significance level of the test then we have

$$EG\left[\frac{\sqrt{\chi'^2_{n_2-1}}}{\sqrt{1+\theta}}\phi\left(\theta\frac{\chi^2_{n_2-1}}{\chi'^2_{n_2-1}}\right)\right] = 1-\alpha,\ 0<\theta<\infty$$

or equivalently

$$EG\left[\frac{\sqrt{Y}}{\sqrt{1+\theta}}\phi\left(\theta\frac{X}{Y}\right)\right] = 1-\alpha,\ 0<\theta<\infty \quad \ldots \quad (9.11)$$

where $\chi^2_{n_1-1} = X$, $\chi^2_{n_2-1} = Y$.

We have assumed that $\phi(\eta)$ has the right-hand derivative $\phi'(0)$. Let us take formally the derivatives of both sides of (9.11) at $\theta = 0$. We get

$$Eg[\phi(0)\sqrt{Y}]\left[\sqrt{Y}\frac{X}{Y}\phi'(0) - \tfrac{1}{2}\sqrt{Y}\phi(0)\right] = 0. \quad \ldots \quad (9.12)$$

If this formal relation is true then it is clear that $\phi'(0) > 0$; otherwise the left-hand side of (9.12) is negative because of $\phi(0) > 0$, $X > 0$, $Y > 0$ with probability 1. Hence to prove (9.8) it is enough to prove (9.12) and this is done in [3, ch. 8] it is enough to prove (9.12) and this is done in [3, ch. 8] for $n_2 \geqslant 4$ by means of rather difficult analytic calculations.

We are now able to achieve the proof of Theorem 9.1. Let $n_2 \geqslant 4$, then Lemma 4 guarantees the existence of a tangency point of the curve $\xi = \phi(\eta)$ with a critical ellipse $B(\xi, \eta) = \mathfrak{A}$, $\mathfrak{A} > 1$. This ellipse will intersect the axis $\xi = 0$ at a point $(0, \eta'_0)$, where $1 \leqslant \eta'_0 \leqslant \frac{3}{2}(M+1)$. According to the conditions of the theorem $\phi(\eta)$ satisfies the Lipschitz condition for $1 < \eta < 2(M+1)$. Then Lemma 2 implies that the test is not similar.

REFERENCES

[1] A. Wald, Testing the difference between the means of two normal populations with unknown standard deviations, *Selected Papers in Prob. and Stat.*, N.Y. McGraw-Hill, 1955, pp. 669–695.

[2] H. Brény, L'état actuel du probleme de Behrens-Fisher (in French), *Trabajos de Estadist.*, 6 (1955), pp. 111–131.

[3] Yu. V. Linnik, Statistical problems with nuisance parameters, Proividence, *Amer. Math Soc.*, 1968,

CHAPTER TEN

CONSTRUCTION OF A NON-RANDOMIZED HOMOGENEOUS SIMILAR TEST FOR THE BEHRENS-FISHER PROBLEM

The results of the previous chapter show us that non-randomized homogeneous similar tests, which have the critical regions of the form

$$G\left(\frac{\bar{x}-\bar{y}}{s_2}, \frac{s_1}{s_2}\right) \geqslant 0 \qquad \ldots \quad (10.1)$$

must be very irregular from the analytic point of view. In fact we have seen that the boundary of the critical region of a Fisher-Welch-Wald test may not be even a differentiable function. These results could lead to the conclusion that in general there are no non-randomized homogeneous similar tests which depend only upon sufficient statistics. However, in 1964 it was proved (see [1, 2] in the monograph [3]) that the similar tests of the form (10.1) exist for any significance level α. We expound here these results [2, 3].

Theorem 10.1. *For any given level $\alpha \in (0, 1)$ and any pair of samples whose sizes n_1 and n_2 are of different parity (i.e. one of them even and the other odd) there exists a non-randomized similar test for the Behrens-Fisher problem whose critical region is defined by the values of $\frac{\bar{x}-\bar{y}}{s_2}$ and $\frac{s_1}{s_2}$. Even a stronger result can be proved.*

Theorem 10.2. *Let k be a finite number of pairs (n_{11}, n_{21}), (n_{12}, n_{22}), $\ldots, (n_{1k}, n_{2k})$ of different parity. Then there exists a non-randomized homogeneous test with the critical region*

$$G\left(\frac{|\bar{x}_i-\bar{y}_i|}{s_{2i}}, \frac{s_{1i}}{s_{2i}}\right) \geqslant 0 \qquad \ldots \quad (10.2)$$

(where $i = 1, \ldots, k$ and the function G is the same for all i) which is similar for all pairs $(n_{11}, n_{21}), \ldots, (n_{1k}, n_{2k})$ of samples and has a level $\alpha \in (0, 1)$ given in advance.

We shall prove Theorem 10.1. To construct a non-randomized test

$$\phi = \phi\left(\frac{|\bar{x}-\bar{y}|}{s_2}, \frac{s_1}{s_2}\right)$$

of level $\alpha \in (0, 1)$ it is sufficient to construct the cotest $\psi = \phi - \alpha$, where the function ψ assumes only the values $-\alpha$ and $(1-\alpha)$. Let us introduce again the variables $\xi = \frac{\bar{x}-\bar{y}}{s_2}$ and $\eta = \frac{s_1}{s_2}$ which we made use of in Chapter eight, and the critical curves $A(\xi, \eta)$, $B(\xi, \eta)$. Let $\phi(\xi, \eta)$ be the indicator function of the critical region of a test sought (i.e. $\phi(\xi, \eta) = 1$ if the point (ξ, η) belongs to the critical region and $\phi(\xi, \eta) = 0$ otherwise) and Ω_1 denote the quadrant $\xi \geqslant 0$, $\eta \geqslant 0$. To determine $\phi(\xi, \eta)$ we return to an equation of the form (8.10). Since θ is positive (we have nothing to do with complex values here), the denominator can be decomposed into the product of two factors :

$$\iint_{\Omega_1} \frac{\eta^{n_1-2}\phi(\xi, \eta)d\xi d\eta}{(\theta+A)^N(\theta+B)^N} = C_{n_1, n_2}\frac{\alpha}{2}\theta^{-\frac{n_1}{2}}(1+\theta)^{-N+\frac{1}{2}}. \quad \ldots \quad (10.3)$$

This relation must hold for all $\theta > 0$; here $N = \frac{n_1+n_2-1}{2}$ and $\alpha \in (0, 1)$ is a given level.

For the cotest $\psi(\xi, \eta)$ we have the relation

$$J(\theta) = \iint_{\Omega_1} \frac{\eta^{n_1-2}\psi(\xi, \eta)d\xi d\eta}{(\theta+A)^N(\theta+B)^N} = 0. \quad \ldots \quad (10.4)$$

If the function $\psi(\xi, \eta)$ assumes only the values $-\alpha$ and $1-\alpha$ and satisfies the relation (10.4) for all $\theta > 0$, it will be the desired cotest. From the assumed difference of parity of n_1, n_2 it follows that N is an integer. Suppose furthermore that $n_1 \geqslant 2$, $n_2 \geqslant 2$; since these numbers are of different parity at least one of them is not less than 3 and $N \geqslant 2$.

In the quadrant Ω_1 : $\xi \geqslant 0$, $\eta \geqslant 0$, the critical curves $A(\xi, \eta)$, $B(\xi, \eta)$ form a system of curvilinear coordinates ; through each point

of the quadrant passes one and only one critical curve of a given type (ellipse or hyperbola). This coordinate system will be convient for the investigation of the integral (10.4). Let as calculate firstly the Jacobian $\frac{\partial(\xi, \eta)}{\partial(A, B)}$ for the points of Ω_1. Since

$$(\theta+A)(\theta+B) = \theta^2+\theta(1+\xi^2+\eta^2)+\eta^2$$

we have

$$\xi = [(B-1)(1-A)]^{\frac{1}{2}}, \quad \eta = (AB)^{\frac{1}{2}}$$

$$\frac{\partial(\xi, \eta)}{\partial(A, B)} = \begin{vmatrix} -\frac{1}{2}(B-1)^{\frac{1}{2}}(1-A)^{-\frac{1}{2}} & \frac{1}{2}(B-A)^{-\frac{1}{2}}(1-A)^{\frac{1}{2}} \\ \frac{1}{2}A^{-\frac{1}{2}}B^{\frac{1}{2}} & \frac{1}{2}A^{\frac{1}{2}}B^{-\frac{1}{2}} \end{vmatrix}$$

$$= -\frac{1}{4}\frac{B-A}{(AB)^{\frac{1}{2}}(B-1)^{\frac{1}{2}}(1-A)^{\frac{1}{2}}}$$

what shows the presence of irregularities on the lines $A = 0$, $B = 0$, $A = 1$, $B = 1$. The point $(A, B) = (1, 1)$ attracts in this sense special attention.

Making use of this expression for the change of coordinates in (10.4) we obtain

$$J(\theta) = \iint_{\Pi} \frac{\psi_1(A, B)(AB)^{\frac{n_1-3}{2}}(B-A)dAdB}{\sqrt{1-A}\sqrt{B-1}(\theta+A)^N(\theta+B)^N} = 0. \quad \ldots \quad (10.5)$$

Here $\psi_1(A, B) = \psi(\xi, \eta)$, π is the half-strip : $0 \leqslant A \leqslant 1$, $B \geqslant 1$, which is the image of the quadrant Ω_1 for the transformation $(\xi, \eta) \to (A, B)$.

Let us partition now the half-strip Π into the disjoint half-strips

$$\Pi_k : 1-2^{-k} \leqslant A < 1-2^{-k-1}, \quad 1 \leqslant B < \infty, \quad k = 0,1,2,\ldots$$

$$\Pi = \bigcup_{k=0}^{\infty} \Pi_k.$$

We shall construct the function $\psi_1(A, B)$ in such a way that for every $k = 0, 1, 2, \ldots$, we have

$$J_k(\theta) = \iint_{\Pi_k} \frac{\psi_1(A, B)(AB)^{\frac{n_1-3}{2}}(B-A)dAdB}{\sqrt{1-A}\sqrt{B-1}(\theta+A)^N(\theta+B)^N} = 0. \quad \ldots \quad (10.6)$$

Assume, without loss of generality, that $n_1 \geqslant 3$. The function $\frac{B-A}{\sqrt{1-A}\,\sqrt{B-1}}$ has only an integrable singularity at the point (1, 1). Since $\frac{n_1-3}{2} \geqslant 0$ it follows from (10.6) ($k = 0, 1, 2, \ldots$) that $J(\theta) = 0$. Since N is an integer and $A \leqslant B$ in the half-strip Π_k we have the decomposition into simple fractions:

$$\frac{1}{(\theta+A)^N(\theta+B)^N} = \sum_{m=1}^{N} \mathfrak{A}_m \frac{(B-A)^{m-1}}{(B-A)^{2N-1}(\theta+A)^m} + \sum_{m=1}^{N} E_m \frac{(B-A)^{m-1}}{(B-A)^{2N-1}(\theta+B)^m}$$

where $\mathfrak{A}_m$, E_m are constants. Substituting this expression into the left-hand side of (10.6) we obtain

$$J_k(\theta) = \sum_{m=1}^{N} \mathfrak{A}_m \int_{1-2^{-k}}^{1-2^{-k-1}} dA \int_1^{\infty} dB \frac{\psi_1(A, B)(AB)^{\frac{n_1-3}{2}}(B-A)^m}{\sqrt{1-A}\,\sqrt{B-1}\,(B-A)^{2N-1}(\theta+A)^m}$$

$$+ \sum_{m=1}^{N} E_m \int_1^{\infty} dB \int_{1-2^{-k}}^{1-2^{-k-1}} dA \frac{\psi_1(A, B)(AB)^{\frac{n_1-3}{2}}(B-A)^m}{\sqrt{1-A}\,\sqrt{B-1}(B-A)^{2N-1}(\theta+B)^m}.$$

Let us partition now the half-strip π_k into the rectangles

$$Q_{ks}: 1-2^{-k} \leqslant A < 1-2^{-k-1}, \quad s \leqslant B < s+1, \quad s = 1, 2, \ldots$$

Consider the functions

$$P_m(A, B) = \frac{(AB)^{\frac{n_1-3}{2}}(B-A)^m}{\sqrt{1-A}\,\sqrt{B-1}\,(B-A)^{2N-1}}, \quad m = 1, 2, \ldots, N.$$

Since $B \geqslant A$ we have $P_m(A, B) \geqslant 0$. For each rectangle Q_{ks} we choose a constant $C_{ks} > 0$ such that the functions

$$p_{mks}(A, B) = \frac{1}{C_{ks}} P_m(A, B), \quad m = 1, 2, \ldots, N$$

will be probability densities with respect to Lebesgue measure. Note also that the functions $p_{mks}(A, B)$ are continuous in the closure $\bar{Q}_{ks}$ of Q_{ks} for $k = 0, 1, 2, \ldots$.

We apply now the I. Romanovskiĭ–V. Rudakov theorem to the densities $p_{mks}(A, B)$ consdered on $\overline{Q}_{ks}$ for each k, s, N.

Theorem (I. Romanovskiĭ-V. Sudakov). *Suppose that on a rectangle* $M : a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$ *there are given* n *probability densities* $p_s(x, y)$, $s = 1, 2, \ldots, n$. *Then for any given* $\alpha \in (0, 1)$ *it is possible to construct a set* $A = A_\alpha \subset M$ *such that*

$$\int_a^b I_A(x, y)p_s(x, y)dx = \alpha \int_a^b p_s(x, y)dx$$

for almost all $y \in [c, d]$, and

$$\int_c^d I_A(x, y)p_s(x, y)dy = \alpha \int_c^d p_s(x, y)dy$$

for almost all $x \in [a, b]$.

Here $I_A(x, y)$ denotes the indicator function of the set A.

The proof of this theorem is based on certain arguments of functional analysis and is rather complicated. Two different proofs can be found in [3, 4].

We are going now to apply the I. Romanovskiĭ-V. Sudakov theorem to the densities $p_{mks}(A, B)$ for all values of k, s, m. For given k, s let us write Q_{ks} in the form

$$a \leqslant A \leqslant b, \quad c \leqslant B \leqslant d.$$

By the I. Romanovskiĭ-V. Sudakov theorem we can construct in the plane a measurable set $\mathcal{U} = \mathcal{U}_{ks}$ such that if $I(A, B) = I_{ks}(A, B)$ denotes its indicator function then

$$\int_a^b I(A, B)p_{mks}(A, B)dA = \alpha \int_a^b p_{mks}(A, B)dA, \; m = 1, 2, \ldots, N$$

for almost all $B \in [c, d]$ and

$$\int_c^d I(A, B)p_{mks}(A, B)dB = \alpha \int_c^d p_{mks}(A, B)dB$$

for almost all $A \in [a, b]$. Here $\alpha \in (0, 1)$ is a level of the desired cotest. Since $I(A, B)$ assumes only the values 0 and 1 the function $\psi_1(A, B) = I(A, B) - \alpha$ assumes only the values $-\alpha$ and $1-\alpha$. Obviously, for $m = 1, 2, \ldots, N$

$$\int_a^b \psi_1(A, B)p_{mks}(A, B)dA = 0 \text{ for almost all } B \in [c, d]$$

and

$$\int_c^d \psi_1(A, B)p_{mks}(A, B)dB = 0 \text{ for almost all } A \in [a, b].$$

Since the function p_{mks} differs from $P_m(A, B)$ only by a constant factor we may define the function $\psi_1(A, B)$ in the interior of the half-strip π by setting it equal to p_{mks} on Q_{ks}, $k = 0, 1, 2, \ldots$; $s = 1, 2, \ldots$. Then we obtain

$$J_k(\theta) = 0, \; k = 1, 2, \ldots, \text{ and } J(\theta) = 0 \text{ for } \theta > 0.$$

This completes the proof of Theorem 10.1.

To prove Theorem 10.2 we consider the finite number of functions $P_{mi}(A, B)$, $i = 1, \ldots, k$; $\ldots = 1, \ldots, N$ (instead of N functions $P_m(A, B)$ in Theorem 10.1):

$$P_{mi}(A, B) = \frac{(AB)^{\frac{n_{1i}-3}{2}}(B-A)^m}{\sqrt{1-A}\sqrt{B-1}(B-A)^{2N_i-1}}$$

where $N_i = \dfrac{n_{1i}+n_{2i}-1}{2}$ and the variables A, B of $P_{mi}(A, B)$ are expressed in terms of $\bar{x}_i$, $\bar{y}_i$, s_{1i}, s_{2i}.

It is unknown whether or not there exist non-randomized homogeneous similar tests when n_1 and n_2 are of equal parity.

REFERENCES

[1] A. M. Kagan, O. V. Shalaevskiĭ, The Behrens-Fisher problem: existence of similar regions in the algebra of sufficient statistics (in Russian), *Doklady Akad. Nauk. SSSR*, 155, 6 (1964).

[2] Yu. V. Linnik, I. V. Romanovskiĭ, V. N. Sudakov, A non-randomized homogeneous test in the Behrens-Fisher problem (in Russian), *Doklady Akad. Nauk SSSR*, 155, 6 (1964).

[3] Yu. V. Linnik, Statistical problems with nuisance parameters. Providence, *Amer. Math. Soc.* 1968.

[4] I. V. Romanovskiĭ, V. N. Sudakov, On the existence of independent partitions (in Russian), *Trudy Mathem. Inst. Steklova Akad. Nauk SSSR*, 79 (1965), pp. 1–7.

CHAPTER ELEVEN

VERIFIABLE FUNCTIONS

We shall discuss here in detail the topic of verifiable functions outlined in Chapter 4. Let us recall the basic definition. Suppose that X is a random variable with a distribution P_θ depending upon an abstract parameter $\theta \in \Omega$. A function $f(\theta)$ is called non-sequentially verifiable if there exists a randomized non-sequential test $\phi(X)$ whose power function $\beta(\theta)$ depends only upon $f(\theta)$, i.e. $\beta(\theta) = \psi(f(\theta))$. A function $f(\theta)$ is called sequentially verifiable if there exists a sequential test ϕ based on independent random variables $X_1, X_2, \ldots$ identically distributed as X such that $E_0\phi = \beta(\theta) = \psi(f(\theta))$. We have already mentioned in Chapter 4 certain results on this topic [1–3]. The results of [3] will be presented later in this chapter. We shall begin with the results of E. L. Lehmann [2] showing the relation between verifiability and distinguishability of the families of distributions.

Two families of distributions

$$P^{(0)} = \{P_\theta;\ \theta \in \omega_0\} \text{ and } P^{(1)} = \{P_\theta;\ \theta \in \omega_1\}$$

where ω_0 and ω_1 are subsets of Ω are called distinguishable if there exists a test $\phi(X)$ with

$$\sup_{\theta \in \omega_1} E_\theta\phi(X) < \inf_{\theta \in \omega_1} E_\theta\phi(X). \qquad \ldots \quad (11.1)$$

Replacing ϕ by $1-\phi$ we see that the definition is symmetric with respect to $P^{(0)}$ and $P^{(1)}$. A very simple geometric interpretation of (11.1) can be given. Denote by α_0 and α_1 the left and right-hand sides of (11.1) respectively and let S be the set of points (α_0, α_1) obtained when ϕ runs over the set of all tests. Then (11.1) means that there exists a point in S which lies above the diagonal formed by the trivial tests $\phi \equiv \alpha$.

Suppose that all the distributions are absolutely continuous with respect to a measure μ and denote by p_0 and p_1 the densities of two distributions chosen in $P^{(0)}$ and $P^{(1)}$ respectively. It is easily seen as in the note by W. Hoeffding [4] that the necessary and sufficient condition of indistinguishability of $P^{(0)}$ and $P^{(1)}$ is the existence for any given $\epsilon > 0$ distributions $P_0 \in P^{(0)}$ and $P_1 \in P^{(1)}$ such that

$$\int |p_1 - p_0| \, d\mu < \epsilon. \qquad \ldots \quad (11.2)$$

It is evident that if a function $f(\theta)$ is verifiable then for any value f_0 there exists a value f_1 such that the set $\omega_i = \{\theta : f(\theta) = f_i\}$, $i = 0, 1$ (or, more precisely the corresponding families of distributions) are distinguishable. It is possible to combine this with (11.2) to obtain a method of studying the phenomenon of indistinguishability.

As an example consider the case of n random variables $X_1, \ldots, X_n$ with the following joint density with respect to Lebesgue measure :

$$\frac{1}{\sigma_n} p\left(\frac{x_1 - \xi}{\sigma}, \ldots, \frac{x_n - \xi}{\sigma} \right), \quad \sigma > 0$$

where p is a continuous function and ξ, σ are unknown parameters. Let $f(\xi, \sigma)$ be a function such that

$$\begin{aligned} f(\xi, \sigma) &< 0 \quad \text{if } \xi < 0, \\ f(\xi, \sigma) &= 0 \quad \text{if } \xi = 0, \\ f(\xi, \sigma) &> 0 \quad \text{if } \xi > 0. \end{aligned}$$

Then the necessary and sufficient condition for distinguishability of the families corresponding to the subsets $\{(\xi, \sigma) : f = 0\}$ and $\{(\xi, \sigma) : f = f_1 > 0\}$ is

$$I = \inf_{f(\xi, \sigma) = f_1} \frac{\xi}{\sigma} > 0. \qquad \ldots \quad (11.3)$$

In fact suppose that $I = 0$ and let (ξ_j, σ_j) be a sequence of points with $f(\xi_j, \sigma_j) = f_1$ and $\frac{\xi_j}{\sigma_j} \to 0$, $j \to \infty$. Let $p_1^{(j)}$ and $p_0^{(j)}$ denote the densities corresponding to (ξ_j, σ_j) and $(0, \sigma_j)$ respectively. Then it is easily seen that

$$\int |p_1^{(j)} - p_0^{(j)}| \, d\mu \to 0, \; j \to \infty$$

and the desired result follows from (11.2).

Consider now the special case when f is of the form $f(\xi, 0) = \frac{\xi}{g(\sigma)}$. The relation $f(\xi, \sigma) = f_1$ is equivalent to $\xi = f_1 g(\sigma)$, f_1 being a constant. The condition (11.3) reduces to

$$\inf_{\sigma>0} \frac{g(\sigma)}{\sigma} > 0. \qquad \ldots \quad (11.4)$$

Suppose furthermore that $g(\sigma) = \sigma^\rho$, ρ being a constant. Then the necessary condition (11.4) for distinguishability can be fulfilled only for $\rho = 1$. Thus we have proved once more the results of [1] for the case of $\rho < 1$. For $\rho = 1$ the condition (11.4) is obviously satisfied. If in this case the observations $X_1, \ldots, X_n$ are independent, i.e.

$$\frac{1}{\sigma^n} p\left(\frac{x_1-\xi}{\sigma}, \ldots, \frac{x_n-\xi}{\sigma}\right) = \prod_1^n \frac{1}{\sigma} p\left(\frac{x_i-\xi}{\sigma}\right)$$

the function $f = \frac{\xi}{\sigma}$ is even non-sequentially verifiable. Indeed let us consider the statistic

$$U = \frac{X_1}{\sqrt{Y_2^2+\ldots+Y_n^2}},$$

where $Y_j = X_j - \overline{X}$, $j \geqslant 2$. Since its distribution depends only upon $\frac{\xi}{\sigma}$ (and depends essentially) the statistic U allows us to construct a non-trivial test whose power function will depend only on $\frac{\xi}{\sigma}$.

We pass now to the method due to V. Palamodov [3] of investigating the non-sequentially verifiable functions in the case of sampling from a normal population. This method introduces in a new manner the theory of functions of one complex variable into mathematical statistics.

Let $X_1, \ldots, X_n$ be independent random variables, $X_i \sim N(a, \sigma^2)$, $i = 1, \ldots, n$. We shall describe all the polynomials $P\left(a, \frac{1}{\sigma^2}\right)$ depending on a and $\frac{1}{\sigma^2}$ admitting invariant verification (the corresponding definition will be modified a little). The method due to V. Palamodov

is purely analytic in contrast with the statistical method due to Lehmann mentioned above.

Without any loss in generality in studying the phenomenon of invariant verification we may restrict ourselves with the tests $\phi(\bar{x}, s)$ depending only on the sufficient statistics

$$\bar{x} = \frac{1}{n} \sum_1^n x_i, \quad s^2 = \frac{1}{n} \sum_1^n (x_i - \bar{x})^2.$$

Setting $\xi = \frac{1}{\sigma^2}$ we shall have

$$E_{a,\sigma}\phi(\bar{x}, s) = \bar{\phi}(a, \sigma^2) = \check{\phi}(a, \xi)$$

$$= C_0 \int_{-\infty}^{\infty} \int_0^{\infty} \xi^{n/2} \exp\left[-\xi(s+(\bar{x}-a)^2]s^{\frac{n-3}{2}}\right. \phi(\bar{x}, s)d\bar{x}ds, \quad \ldots \quad (11.5)$$

C_0 being a constant. Note that the integral (11.5) can be continued as an analytic function of two complex variables to the product of the complex C of a-values and the complex halfplane C^+ of ξ-values with $\mathrm{Re}\,\xi > 0$. This enables us to introduce the following definition.

Definition. A function $f(a, \xi)$ defined on $C \times C^+$ is called C-verifiable if there exists a test ϕ such that

$$\check{\phi}(a, \xi) = \psi(f(a, \xi)), \; a \in C, \quad \xi \in C^+$$

for some $\psi \equiv \text{const}$.

Any real C-verifiable function is obviously verifiable in the sense of the definition given in Chapter 4. The converse is not true, see [3].

In [3] V. Palamodov starts with the proof of the following lemma.

Lemma 1. For any test $\phi(\bar{x}, s)$

$$|\check{\phi}(a, \xi)| \leqslant C_0 \left|\frac{\xi}{\mathrm{Re}\,\xi}\right|^{n/2} \exp\left[\frac{|\xi|^2}{\mathrm{Re}\,\xi} |I_m a|^2\right] \quad \ldots \quad (11.6)$$

for $a \in C$, $\xi \in C^+$ (C_0, C_1 in what follows are positive constants).

The proof of the pure analytic inequality is based on the fact that $0 \leqslant \phi(\bar{x}, s) \leqslant 1$; we shall omit it referring the reader to [3].

We say that $r(a, \xi) \in R$ if

$$r(a, \xi) = \frac{p(a, \xi)}{q(a, \xi)} = \frac{p_m(\xi)a^m + \ldots + p_0(\xi)}{q_k(\xi)a^k + \ldots + q_0(\xi)} \qquad \ldots \quad (11.7)$$

where p_i, q_j $(i = 0, 1, \ldots, m;\ j = 0, 1, \ldots, k)$ belong to the ring $\mathcal{A}$ of the functions holomorphic in C^+ and $p_m q_k \neq 0$. Without loss of generality we suppose that either $m > k$ or $m = k$ but $\frac{p_m(\xi)}{q_k(\xi)} \neq \text{const.}$ Moreover we suppose that the number m is the least possible and the functions p_i and q_j do not vanish simultaneously.

Theorem 11.1. (V. Palamodov). 1. *If a function $r \in R$ with $m > 0$ is C-verifiable, then $k = 0$ in* (11.7) *and $q_0(\xi) \neq 0$ on C^+. Moreover, the power function of the test ϕ with $\check{\phi}(a, \xi) = \psi(r(a, \xi))$ is an entire function of an order not exceeding $\frac{2}{m}$.*

2. *If, moreover, the functions $q_0, p_1, \ldots, p_m$ and $\frac{p_0}{q_0}$ are analytic in the vicinity of zero, the C-verifiable function is necessarily of the form*

$$r(a, \xi) = r_2(\xi)a^2 + r_1(\xi)a + r_0(\xi), \quad r_i = \frac{p_i}{q_0} \in \mathcal{A}$$

upto multiplication by a constant. We also have $r_2(0) = r_1(0) = 0$, $r_2'(0) = 1$ and the image of the mapping $r_2 : C^+ \to C$ belongs to C^+. The function ψ is bounded in any closed angle belonging to C^+.

In what follows we shall denote by C^* the complex plane compactificated in the usual way.

Another purely analytic lemma whose proof is given in [3] is the following.

Lemma 2. Let $\rho : \omega \to C^*$ can be non-constant function meromorphic in a domain ω and Ω be the range of ρ. If $\psi(\rho(z))$, $z \in \omega$, is analytic in ω then ψ is analytic in Ω.

We pass now to the proof of Theorem 11.1. Suppose that $\psi \neq \text{const.}$ It is easy to see that the range of the function $r : C \times C^+ \to C^*$ always contains C; hence by Lemma 2, ψ is an entire function.

If $k > 0$, then for a suitable ξ_0 the image of the mapping $r(a, \xi_0) : C \to C^*$ contains the point ∞. By virtue of Lemma 2 the function

ψ is analytic at that point. Since it is an entire function it must be a constant in contradiction with our assumption. Hence $k = 0$. In a similar way we can prove that q_0 has no zeros in C^+ which implies the relation

$$r(a, \xi) = r_m(\xi)a^m + \ldots + r_0(\xi), \; r_i = \frac{p_i}{q_0} \in \mathcal{A}.$$

Let us fix a point ξ so that $r_m(\xi) \neq 0$. Then $r(a, \xi) \sim r_m(\xi)a^m$ for $|a| \to \infty$. The indequality (11.6) implies for certain values of the constants A and B that

$$|\psi(r(a, \xi))| \leqslant C_0 \exp A |I_m a|^2 \leqslant \leqslant C_0 \exp B |r(a, \xi)|^{2/m}. \quad \ldots \quad (11.8)$$

Thus the order of the entire function ψ does not exceed $\frac{2}{m}$. This proves the first assertion of Theorem 11.1

We pass now to the proof of the second assertion. The functions $r_0, p_1, \ldots, p_m, q_0$ are analytic in the vicinity of zero. Hence the functions $r_i = \frac{p_i}{q_0}$ are analytic in the vicinity of zero, except perhaps at zero itself, where the only possible singularity is a pole. We shall show that in fact the coefficients r_i, $i = 1, \ldots, m$ are analytic in the vicinity of zero and $r_i(0) = 0$. For each $i = 1, \ldots, m$ we have in the vicinity of zero

$$r_i(\xi) = \rho_i \xi^{\alpha_i} + \rho_i' \xi^{\alpha_{i+1}} + \ldots$$

which contain $\rho_i \neq 0$ and α_i. If all $\alpha_i > 0$ the desired assertion is proved. Suppose that $\alpha_i < 0$ for a certain i. Consider the quantity

$$\alpha = \max_{i \geqslant 1} \left(-\frac{\alpha_i}{i}\right).$$

Let k be the largest number for which $-\frac{\alpha_k}{k} = \alpha$. For $\lambda \in C$ set

$$a_\lambda(\xi) = \left[\frac{\lambda}{r_m(\xi)}\right]^{1/n}, \quad \xi > 0$$

where the branch of the root may be chosen arbitrarily. Since $\alpha_k < 0$, $a_\lambda(\xi)$ is bounded when $\xi \to 0$ for any $\lambda \in C$. From $\xi > 0$ and (11.7) we obtain

$$|\psi(r(a_\lambda(\xi), \xi))| \leqslant C \exp B(\lambda)\xi. \qquad \ldots \quad (11.9)$$

On the other hand

$$r(a_\lambda(\xi), \xi) = \sum_{j>k} r_j(\xi) a_\lambda(\xi)^j + \lambda + \sum_{j<k} r_j(\xi) a_\lambda(\xi)^j. \qquad \ldots \quad (11.10)$$

We have for the first sum of the right-hand side

$$|r_j(\xi) a_\lambda(\xi)^j| = \left[\frac{|r_j(\xi)|^{1/j}}{|r_k(\xi)|^{1/k}}\right] |\lambda|^{j/k} \sim C\xi^{j\left(\frac{\alpha_j}{j} - \frac{\alpha_k}{k}\right)} \to 0$$

for $\xi \to 0$ because of the choice of k. For the second sum we have

$$|r_j(\xi) a_\lambda(\xi)^j| \leqslant C\xi^{j\left(\frac{\alpha_j}{j} - \frac{\alpha_k}{k}\right)} |\lambda|^{j/k} \leqslant C\left(|\lambda|^{\frac{k-1}{k}} + 1\right)$$

for $0 < \xi < 1$, since $\frac{\alpha_j}{j} \geqslant \frac{\alpha_k}{k}$ for $j < k$ and the function $r_0(\xi)$ is bounded when $\xi \to 0$. We choose ξ so small that the first sum in (11.10) is smaller than 1. In this case

$$|r(a_\lambda(\xi), \xi) - \lambda| \leqslant C\left(|\lambda|^{\frac{k-1}{k}} + 1\right) \qquad \ldots \quad (11.11)$$

We take now a sufficiently large R. If the point λ runs over the circumference of radius R, the inequality (11.11) implies that the point $r(a_\lambda(\xi), \xi)$, which depends on λ continuously, runs over a curve homeomorphic to that circumference and containing the circle

$$|z| \leqslant R' = R - C\left(R^{\frac{k-1}{k}} + 1\right).$$

On the other hand from (11.9) it follows that the function ψ is bounded by the constant

$$C \exp B(\lambda)\xi.$$

Since ξ can be chosen arbitrarily small, ψ is bounded by the constant C, which does not depend upon λ on the curve under consideration and therefore in the circle of radius R'. As $R \to \infty$ so does R'; hence

ψ is bounded on the whole plane which leads to $\psi \equiv \text{const}$. Since we have assumed that $\psi \neq \text{const}$ we see that all $\alpha_i > 0$.

Let us prove now the next lemma.

Lemma 3. For any $\xi \in C^+$ there exists an $\epsilon > 0$ such that the function ψ is bounded inside the angle

$$|\arg z - \arg r_m(\xi)| \leqslant \epsilon \qquad \ldots \quad (11.12)$$

and, for odd m, also inside the angle

$$|\arg z - \arg(-r_m(\xi))| \leqslant \epsilon.$$

To prove Lemma 3 let us take an arbitrary $\xi \in C^+$. Since $r_m(0) = 0$, it follows that $r_m(\xi) \not\equiv \text{const}$. Therefore we can find in the vicinity of the point ξ two points ξ_+ and ξ_- such that

$$\arg r_m(\xi_+) > \arg r_m(\xi) > \arg r_m(\xi_-)$$

$$\arg r_m(\xi_+) - \arg r_m(\xi_-) = \delta > \frac{\pi}{2} \qquad \ldots \quad (11.13)$$

We shall show now that the function ψ is bounded inside the angle

$$\arg r_m(\xi_+) - \frac{\delta}{3} \geqslant \arg z \geqslant \arg r_m(\xi_-) - \frac{\delta}{3}. \qquad \ldots \quad (11.14)$$

Consider the curves

$$\theta_\pm(a) = r(a, \xi_\pm), \quad a \in [0, \infty). \qquad \ldots \quad (11.15)$$

We have

$$\theta_\pm(a) - r_m(\xi_\pm) a^m = 0(a^{m-1}), \quad a \to \infty.$$

This means that the domain between the curves $\theta_\pm(a)$, taken together with a sufficiently large circle, contains the angle (11.14).

According to the inequality (11.6) the function $\psi(r(a, \xi_\pm))$ is bounded for real values of a and therefore the function ψ is bounded on the curves (11.15). But by (11.13) these curves are contained inside an angle less than $\frac{\pi}{2}$ while ψ was shown above to be an entire function of the order not exceeding $\frac{2}{m} \leqslant 2$.

Hence by the Phragmén-Lindelöf principle, the function ψ is bounded in the domain lying between the curves (11.15) as was to be proved.

The case of an odd m is dealt with by analogy with ξ replaced by $-\xi$.

To complete the proof of Theorem 11.1 consider the set

$$\{\arg r_m(\xi),\ \xi \in C^+\}. \qquad \ldots \quad (11.16)$$

Since

$$r_m(\xi) = \rho_m \xi^{\alpha_m} + \rho_m' \xi^{\alpha_m+1} + \ldots, \quad \alpha_m > 0$$

for sufficiently small ξ, we have

$$\arg r_m(\xi) \sim \alpha_m \arg \xi + \arg \rho_m, \quad \xi \to 0.$$

Therefore if $\alpha_m > 1$ the set (11.16) contains an interval of length 2π. Lemma 3 implies then that ψ is bounded inside the angle $2\pi - \epsilon$ for any $\epsilon > 0$. As ψ is an entire function of a finite order, we have by the Phragmén-Lindelöf principle that $\psi \equiv \text{const}$. Hence $\alpha_m = 1$. Multiplying $r(a, \xi)$ by $\frac{1}{p_m}$ we get $r_m'(0) = 1$. Then (11.17) shows that the set (11.16) contains the interval $\left(-\frac{\pi}{2}, \frac{\pi}{\pi}\right)$ and by Lemma 3 the function ψ is bounded inside the angle $|\arg z| \leqslant \frac{\pi}{2} - \epsilon$ for any $\epsilon > 0$. Hence ψ is a function of order at least 1 and of finite type. On the other hand we have already proved that its order does not exceed $\frac{2}{m}$. Hence $m \leqslant 2$. The case of $m = 1$ is excluded because for $m = 1$ Lemma 3 would imply the boundedness of ψ also in the angle $|\arg z - \pi| \leqslant \frac{\pi}{2} - \epsilon$ for any $\epsilon > 0$ which is impossible for non-constant functions of finite order.

It remains to verify that the image of the mapping $r_2 : C^+ \to C$ belongs to C^+. If it is not so, then by virtue of Lemma 3 the function ψ is bounded inside the angle

$$|\arg z - \arg r_2(\xi)| \leqslant \epsilon,$$

Since ψ is of the first order and bounded inside any angle of the form

$$|\arg z| \leqslant \frac{\pi}{2} - \epsilon$$

again the Phragmén-Lindelöf principle implies that it vanishes identically. This proves Theorem 11.1

REFERENCES

[1] Yu. V. Linnik, Statistical problems with nuisance parameters, Providence, *Amer. Math Soc.*, 1968.

[2] E. L. Lehmann, On the non-verifiability of certain parametric functions, *Teoria veroiatn. i primen.*, **10**, 4 (1965).

[3] V. P. Palamodov, On verifiable functions (in Russian), *Teoria veroiatn. i primen.*, **13**, 1 (1968).

[4] W. Hoeffding, *The role of assumptions in statistical decisions*, Proc. 3rd Berkeley Symposium on Math. Statistics and Probability, Berkeley and Los Angeles, Univ. California Press, Vol. I, 1956.

CHAPTER TWELVE

VERIFIABLE POLYNOMIALS. SEQUENTIALLY VERIFIABLE FUNCTIONS

For the non-sequentially verifiable polynomials V. Palamodov [1] obtained the complete classification.

We shall always consider a sample $x_1, \ldots, x_n$ from a normal population $N(a, \sigma^2)$ and set $\xi = \frac{1}{\sigma^2}$. The classification of the verifiable polynomials $p(a, \xi)$ is given by the following theorem.

Theorem 12.1. *For a polynomial $p(a, \xi)$ which is not a function of ξ only to be C-verifiable it is necessary and sufficient that*

$$p(a, \xi) = \xi(a^2 + Aa + B) \qquad \ldots \quad (12.1)$$

where A and B are real and

$$A^2 - 4B \leqslant 0. \qquad \ldots \quad (12.2)$$

Note that in Chapter 11 we proved that the function $f = \sigma = \xi^{-\frac{1}{2}}$ is verifiable where from it follows the verifiability of any continuous monotone function of ξ.

We shall start with the proof of necessity of the representation (12.1). Let $p(a, \xi)$ be a C-verifiable polynomial and let ϕ be a test such that $\check{\phi}(a, \xi) = \psi(p(a, \xi))$ where $\psi \not\equiv \text{const}$. By Theorem 11.1 ψ is an entire function and the polynomial $p(a, \xi)$ after multiplication by a suitable constant takes on the form

$$p(a, \xi) = p_2(\xi)a^2 + p_1(\xi)a + p_0(\xi) \qquad \ldots \quad (12.3)$$

where

$$p_2(0) = p_1(0), \quad p_2'(0) = 1.$$

The functions $p_2(\xi)$, $p_1(\xi)$, $p_0(\xi)$ are polynomials. We shall show that their degrees are at most 1. Suppose this is not true; then for

a certain $a = a_0$ the polynomial $p(a_0, \xi)$ is of degree $k > 1$ in ξ. Hence

$$p(a_0, \xi) = C\xi^k + 0(|\xi|^{k-1}), \quad |\xi| \to \infty. \qquad \ldots \quad (12.4)$$

If the point ξ is inside the angle $|\arc z| \leqslant \frac{\pi}{2} - \epsilon$ then by virtue of (11.6) we get that $\psi(p(a_0, \xi))$ is bounded. On the other hand, the first term in (12.4) runs over all the values in the angle

$$|\arc z| \leqslant \pi - 2\epsilon.$$

Hence the function $p(a_0, \xi)$ runs over all the values of the complex plane, except may be the values in a certain circle and in the angle

$$|\arc z - \pi| \leqslant 3\epsilon$$

which is as small as we please. Since ψ is a function of finite order, the Phragmén-Lindelöf principle implies that it is bounded on the whole plane and therefore reduces to a constant. This contradiction proves that $p_2(\xi)$, $p_1(\xi)$ and $p_0(\xi)$ are linear functions of ξ.

From (12.3) we have then $p_2(\xi) = \xi$, $p_1(\xi) = A\xi$, where A is a constant. Subtracting a suitable constant from $p(a, \xi)$ we can also annul the constant term of $p_0(\xi)$. Then $p(a, \xi)$ will take on the form (12.3).

Consider now the function $\arc(a^2 + Aa + B)$ for real values of a. Suppose it to be non-constant and let ϕ_1, ϕ_2 be two distinct values of it. In this case if ξ runs over the angle $|\arc \xi| \leqslant \frac{\pi}{2} - \epsilon$ and the point a runs over the real axis, the point $\xi(a^2 + Aa + B)$ will take on all the values inside the angles

$$|\arc z - \phi_1| \leqslant \frac{\pi}{2} - \epsilon \quad \text{and} \quad |\arc z - \phi_2| \leqslant \frac{\pi}{2} - \epsilon.$$

The inequality (11.6) implies that ψ is bounded inside these angles for any $\epsilon > 0$. Since $\phi_1 \neq \phi_2$ and ψ is an entire function, this is impossible in view of the Phragmén-Lindelöf principle. Hence

$$\arc(a^2 + Aa + B) = \text{const.}$$

On the other hand, $\arc(a^2 + Aa + B) = 0$ for $a \to \infty$. Hence A and B are real and $a^2 + Aa + B$ is non-negative for all real a, so that $A^2 - 4B \leqslant 0$.

We pass now to the proof of sufficiency. For a given polynomial of the form (12.1) we shall construct a test ϕ such that

$$\check{\phi}(a, \xi) = C \exp -\tau\xi(a^2+Aa+B). \qquad \ldots \quad (12.5)$$

To this end in view we consider the test

$$\phi(\bar{x}, s) = \begin{cases} \left(1-\frac{t\bar{x}^2}{s}\right)^{\frac{n-3}{2}}, & s \geqslant tx^2 \\ 0, & s < tx^2 \end{cases}$$

where t is an arbitrary number. We shall find now its power function. We have

$$\check{\phi}(a, \xi) = C_n \xi^{n/2} \exp(-\xi a^2) \iint_{S \geqslant t\bar{x}^2} \exp\left[\xi(-s-\bar{x}^2+2a\bar{x})\right](s-t\bar{x}^2)^{\frac{n-3}{2}} \, ds dx.$$

The transformation

$$s \to s' = s-t\bar{x}^2, \quad \bar{x} \to x$$

transfers the right-hand side of the last formula to

$$C_n \xi^{n/2} \exp(-\xi a^2) \int_{-\infty}^{\infty} \int_{0}^{\infty} \exp[-\xi s'-\xi(1+t)x^2+2\xi ax] s'^{\frac{n-3}{2}} \, ds' dx$$

$$= C_n \exp\left(-\xi a^2+\frac{a^2}{1+t}\right) \cdot \xi^{n/2} \int_{-\infty}^{\infty} \int_{0}^{\infty} \exp\left[-\xi s'-\xi\left(\sqrt{1+t}x-\frac{\xi}{\sqrt{1+t}}\right)^2\right]$$

$$\times s'^{\frac{n-3}{2}} \, ds' dx = C_n \exp(-\tau\xi a^2),$$

where $\tau = 1-\frac{1}{1+t}$. The power function of the test $\phi(\bar{x}-y, s)$, y beying an ancillary number, is $C \exp(-\tau\xi(a-y)^2)$. The formula (11.5) gives now the following expression for the power function of the test $\phi(\bar{x}-y)$, $s-s_0)$, where s_0 is a constant :

$$C \exp(-\tau\xi(a-y)^2 \exp(-s_0\xi) = C \exp\left[-\tau\xi(a-y)^2+\frac{s_0}{\tau}\right].$$

By a suitable choice of the constants y and s_0 we can transform the last expression to the form (12.5).

Till now we have considered the problems of C-verifiability of functions, whereas the verifiability denfied in Chapter 4 (we shall

call it R-verifiability) is of interest from the statistical point of view. The relation between these two types of verifiability was investigated in detail in the paper by V. Palamodov [1]. Though his results are simply formulated, their proofs are profound and complicated and we are not able to reproduce them here, referring the reader to [1].

The principal result by V. Palamodov concerning the R-verifiable functions $r(a, \xi)$ of the form (11.7) is following.

Theorem 12.2. *Let m be positive in the irreducible representation (11.7) of a function $r \in R$. Suppose there exists a non-discrete set $\Lambda \subset C$ with the following property: for all $\lambda \in \Lambda$ any connected component of the curve*

$$r(\xi, a) = \lambda, \quad a \in C, \quad \xi \in C^+$$

has non-empty intersection with $R \times R^+$.

Then if the function r is R-verifiable it will be also C-verifiable.

This theorem has a rather interesting corollary.

Let $r \in R$ be a R-verifiable function. Then if for a point $\xi \in R^+$ it takes on a value λ_0 at exactly m points, m being the exponent in the numerator of the irreducible representation (11.7) of r, r will be C-verifiable. This condition is always fulfilled for $m = 1$ or $m = 2$.

Therefore the second order polynomials with respect to a which are R-verifiable are of the form (12.1).

Let us pass now to sequentially verifiable functions. We shall show that if $x_1, x_2, \ldots$ are independent normally distributed $N(a, \sigma^2)$ variables the function $\gamma(a, \sigma) = a$ is sequentially verifibable. We follow the well known paper by Stein [2].

Suppose that on the base of independent normal $N(a, \sigma^2)$ observations $x_1, x_2, \ldots$, we are going to construct a non-trivial test of the hypothesis $H_0: a = a_0$ whose power function depends only upon $a - a_0$. For a sample $x_1, \ldots, x_{n_0}$ of size n_0 we consider the unbiased estimator of the variance

$$s^2 = \frac{1}{n_0 - 1}\left[\sum_1^{n_0} x_i^2 - \frac{1}{n_0}\left(\sum_1^{n_0} x_i\right)^2\right]. \qquad \ldots \quad (12.6)$$

Let us define the integer n by

$$n = \max \left\{ \left[\frac{s^2}{z} \right] + 1, \quad n_0 + 1 \right\} \qquad \ldots \quad (12.7)$$

where z is a constant given in advance, $[q]$ denotes the greatest integer not exceeding q. Having calculated s^2, we can determine the real numbers b_i, $i = 1, \ldots, n$ such that

$$\sum_1^n b_i = 1, \quad b_1 = b_2 = \ldots = b_{n_0}, \quad s^2 \sum_1^n b_i^2 = z. \qquad \ldots \quad (12.8)$$

Such a choice is possible since, when

$$\sum_1^n b_i = 1, \quad b_1 = b_2 = \ldots = b_{n_0}$$

$\min \sum_1^n b_i^2$ is attained for $b_i = \frac{1}{n}$, $i = 1, \ldots, n$ and its value is $\frac{1}{n}$. From (12.7) we have $n \geqslant \frac{s^2}{z}$ and $\frac{s^2}{n} \leqslant z$.

Introduce now the random variable :

$$t' = \frac{\sum_0^n b_i x_i - a_0}{\sqrt{z}} = \frac{\sum_1^n b_i (x_i - a)}{\sqrt{z}} + \frac{a - a_0}{\sqrt{z}} = u + \frac{a - a_0}{\sqrt{z}} \qquad \ldots \quad (12.9)$$

where

$$u = \frac{\sum_1^n b_i (x_i - a)}{\sqrt{z}}. \qquad \ldots \quad (12.10)$$

Let us show that the random variable u follows the Student distribution with $n_0 - 1$ degrees of freedom (for the definition see, for example, [3]). By virtue of (12.8) we have

$$u = \frac{\sum_1^n b_i (x_i - a)}{s \sqrt{\sum_1^n b_i^2}} = \frac{b_1 \sum_1^{n_0} (x_i - a) + \sum_{n_0+1}^n b_j (x_j - a)}{s \sqrt{\sum_1^n b_i^2}}.$$

Note that the random variable $b_1 \sum_1^{n_0} (x_i - a)$ is independent of s (see Chapter 2). The random variable $\sum_{n_0+1}^n b_j (x_j - a)$ is independent both of s and of $\sum_1^{n_0} (x_i - a)$ because it contains only the observations x_j with $j > n_0$.

Thus for given s and given constants b_i, the numerator of (12.10) is distributed independently of s and follows a normal distribution $N\left(0, \sigma^2 \sum_1^n b_i^2\right)$. The random variable $\frac{(n_0-1)s^2}{\sigma^2}$ is distributed identically with χ_{n_0-1}, where $\chi^2_{n_0-1}$ follows the chi-square distribution with n_0-1 degrees of freedom, so that s is distributed as

$$\sigma \sqrt{\frac{\chi^2_{n_0-1}}{n_0-1}}.$$

Thus the random variable u is distributed as

$$\frac{X}{\sqrt{\frac{\chi^2_{n_0-1}}{n_0-1}}}$$

where $X \sim N(0, 1)$ and X is independent of $\chi^2_{n_0-1}$. Therefore u is distributed identically with t_{n_0-1}, where t_{n_0-1} follows the Student distribution with (n_0-1) degrees of freedom. With the aid of u it is possible to construct a non-randomized sequential test of the hypothesis $H_0 : a = a_0$. Let α be a given level, $\xi_{\alpha/2}$ is determined by the relation

$$P\{t_{n_0-1} > \xi_{\alpha/2}\} = \frac{\alpha}{2}.$$

Let us define the critical region

$$\left| \frac{\sum_1^n b_i x_i - a_0}{\sqrt{z}} \right| > \xi_{\alpha/2}.$$

The power function of the test with this critical region is

$$P\left\{ \left| t_{n_0-1} + \frac{a-a_0}{\sqrt{z}} \right| > \xi_{\alpha/2} \right\},$$

z being a constant given in advance. We see that the power function is independent of σ and depends only on $a-a_0$.

Thus the function $\gamma(a, \sigma) = a$ and any monotone function of a are sequentially verifiable. Other examples of sequentially verifiable functions which can not be verified non-sequentially are not known.

REFERENCES

[1] V. P. Palamodov, On verifiable functions (in Russian), *Teoria veroiatn. i primen.*, **13**, 1 (1968).

[2] C. Stein, A two sample test for a linear hypothesis whose power is independent of the vairance, *Ann. Math. Stat.*, **16**, 2 (1945).

[3] E. L. Lehmann, *Testing Statistical Hypotheses*, N.Y., Wiley, 1959.

CHAPTER THIRTEEN

ON UNBIASED ESTIMATION FOR INCOMPLETE EXPONENTIAL FAMILIES WITH POLYNOMIAL RELATIONS

The classification and optimization problems for unbiased estimators are closely connected with those for similar tests. Indeed suppose we have a family $\{P_\theta\}$ of distributions and let U be an unbiased estimator of a given (scalar-valued) function $\gamma(\theta)$, i.e.

$$E_\theta U = \gamma(\theta) \text{ for all } \theta.$$

Then any other unbiased estimator U_1 or $\gamma(\theta)$ can be written as

$$U_1 = U + \chi$$

where $E_\theta \chi = 0$ for all θ. We shall call χ an unbiased estimator of zero. Obviously an unbiased estimator of zero behaves like a precotest defined in Chapter 5. That is why the study of unbiased estimators can be carried out in parallel with that of precotests and tests and suitable tools are the analytic sheaves and theorems by L. Hörmander. We shall expound here the results of A. Kagan and V. Palamodov on this topic.

We consider the problem of unbiased estimation of parametric functions on the basis of n independent observations $x_1, \ldots, x_n$ with the marginal density

$$f(x; \alpha) = \exp\{t_0(x) + c_1(\alpha)t_1(x) + \ldots + c_s(\alpha)t_s(x) + c_0(\alpha)\} \qquad \ldots (13.1)$$

with respect to the Lebesgue measure.

Here $s \leqslant n$ and α is an abstract parameter, $\alpha \in A$. Introduce the natural parameters

$$\theta_i = c_i(\alpha), \quad i = 1, \ldots, s$$

and suppose that when α runs over A the point $\theta = (\theta_1, \ldots, \theta_s)$ runs over an everywhere dense subset of an algebraic subvariety of the domain $\Omega \subset R^s$. We shall write this subvariety in the form $\Omega \bigcap \Pi$, where Π is an algebraic variety in R^s given by the polynomial relations

$$\begin{aligned} &\Pi_1(\theta_1, \ldots, \theta_s) = 0 \\ &\ldots \quad \ldots \quad \ldots \\ &\Pi_r(\theta_1, \ldots, \theta_s) = 0 \end{aligned} \qquad \ldots \quad (13.2)$$

with $r < s$.

The distribution of the sample $(x_1, \ldots, x_n) = x$ from the population (13.1) is given in R^n by the density

$$f^{(n)}(x; \theta) = [C(\theta)]^n \exp\left\{ \sum_{i=1}^{n} t_0(x_i) + \theta_1 \sum_{i=1}^{n} t_1(x_i) + \ldots + \theta_s \sum_{i=1}^{n} t_s(x_i) \right\} \qquad \ldots \quad (13.3)$$

with respect to the Lebesgue measure, where $C(\theta)$ is determined by the norming condition. The sufficient statistics for the family of densities (13.3) are

$$T_1 = \sum_{i=1}^{n} t_1(x_i), \ldots, \quad T_s = \sum_{i=1}^{n} t_s(x_i).$$

Suppose that $T_1, \ldots, T_s$ are functionally independent, then the distribution of the vector $T = (T_1, \ldots, T_s)$ is given in R^s by the density

$$p(T, \theta) = C(\theta) h(T) \exp(\theta_1 T_1 + \ldots + \theta_s T_s) \qquad \ldots \quad (13.4)$$

with respect to the Lebesgue measure; here $h(T) \geqslant 0$, $\theta \in \Omega \bigcap \Pi$. As before denote by supp h the support of $h(T)$ and by $\mathcal{T}$ its interior ($\mathcal{T} = \text{int supp } h$). We impose on $h(T)$ the following condition

$$\left.\begin{aligned} &h(T) \text{ is infinitely differentiable on } \mathcal{T} \\ &\text{mes supp } h = \text{mes } \mathcal{T}. \end{aligned}\right\} \qquad \ldots \quad (13.5)$$

where mes denotes Lebesgue measure.

Let $g(T)$ be an unbiased estimator of a parameteric function $\gamma(\theta)$ depending only upon the vector T of sufficient statistics and having a finite variance for all values of $\theta \in \Omega \bigcap \Pi$.

We shall investigate here the conditions which are imposed on the variety Π and the estimator $g(T)$ itself by the optimality property

of $g(T)$ for all $\theta \in \Omega \bigcap \Pi$ within the class of all unbiased estimators of $\gamma(\theta)$ with finite variances. Throughout the lecture we take the variance as a measure of quality of an estimator.

The conditions of optimality mentioned above were completely investigated by A. Kagan and V. Palamodov [1]; we expound here their results.

It is clear that the behaviour of $g(T)$ outside $\mathcal{T}$ is of no importance for its property as an estimator of $\gamma(\theta)$; therefore we shall suppose $g(T)$ to be defined only on $\mathcal{T}$.

Let us denote by $\mathcal{N}$ the least complex algebraic variety in C^s containing $\Omega \bigcap \Pi$. Since Π is itself an algebraic variety $\Pi \bigcap \Omega = \mathcal{N} \bigcap \Omega$.

Theorem 13.1. *In order that the function* $g(T)$, $T \in \mathcal{T}$ *for which* $E_\theta g^2 < \infty$; $\theta \in \Omega \bigcap \Pi$ *be the best unbiased estimator of the function* $E_\theta g = \gamma(\theta)$ *for the exponential family* (13.3), $\theta \in \Omega \bigcap \Pi$, *under condition* (13.5) *it is necessary and sufficient that*:

1) *In the space* C^s *of variables* $\theta_1, \ldots, \theta_s$ *there exists a linear system of coordinates* $\theta'_1, \ldots, \theta'_s$ *in which* $\mathcal{N}$ *is a cylinder of type* $L \times \nu$, *where* L *is the coordinate space* $\theta'_1 = \ldots = \theta'_m = 0$ *and* ν *is a certain set in the space*: $\theta'_{m+1} = \ldots = \theta'_m = 0$, $0 \leqslant m \leqslant s$.

2) *In the corresponding system of coordinates* $T'_1, \ldots, T'_s$ *in the space* R^s *of values* $(T_1, \ldots, T_s)$ *the function* $g(T)$ *depends only on* $T'_{m+1}, \ldots, T'_s$.

At first we shall prove the following lemma.

Lemma 1. For $g(T)$ to be the optimal unbiased estimator of $\gamma(\theta)$, $\theta \in \Omega \bigcap \Pi$ it is necessary and sufficient that for each unbiased estimator χ of zero with a finite variance

$$E_\theta(g\chi) = 0, \quad \theta \in \Omega \ \Pi.$$

In fact, let $g(T)$ be the best unbiased estimator of $\gamma(\theta)$ and $\chi(T)$ an arbitrary unbiased estimator of zero. Then for an arbitrary constant c

$$\mathfrak{D}_\theta(g+c\chi) = \mathfrak{D}_\theta g + 2cE_\theta(g\chi) + c^2\mathfrak{D}_\theta\chi. \qquad \ldots \quad (13.6)$$

Since for all values of c

$$\mathfrak{D}_\theta(g+c\chi) \geqslant \mathfrak{D}_\theta g$$

it is easy to deduce from (13.6) that $E_\theta g(\chi) = 0$, $\theta \in \Omega \bigcap \Pi$. Conversely, let $E_\theta(g\chi) = 0$, $\theta \in \Omega \bigcap \Pi$ for any unbiased estimator of zero $\chi(T)$. If the estimator $g_1(T)$ is such that

$$E_\theta g_1 = \gamma(\theta), \quad E_\theta g_1^2 < \infty, \quad \theta \in \Omega \bigcap \Pi$$

then $\chi = g_1 - g$ will be an unbiased estimator of zero. We have:

$$\mathfrak{D}_\theta g_1 = E_\theta(g - \gamma(\theta) + \chi)^2 = \mathfrak{D}_\theta g + \mathfrak{D}_\theta \chi \geqslant \mathfrak{D}_\theta g$$

which proves Lemma 1.

We can now prove the sufficiency of the conditions of Theorem 13.1. Without loss of generality, we can assume that in the initial system itself the subspace L is given by the equations $\theta_1 = \ldots = \theta_m = 0$, and ν is a subset of the subspace $\theta_{m+1} = \ldots = \theta_s = 0$, and that the function $g(T)$ depends only upon $T_{m+1}, \ldots, T_s$. Let $\chi(T)$ be an arbitrary unbiased estimator of zero with $E_\theta \chi^2 < \infty$, $\theta \in \Omega \bigcap \Pi$. Take any point $\theta_0 = (\theta_{10}, \ldots, \theta_{s0}) \in \mathcal{N} \bigcap \Omega$, denote by ω the section of Ω by the surface $\theta_1 = \theta_{10}, \ldots, \theta_m = \theta_{m0}$ and put

$$\psi(T) = \psi(T_{m+1}, \ldots, T_s)$$
$$= \int \chi(T) h(T) \exp(\theta_{10} T_1 + \ldots + \theta_{m0} T_m) dT_1, \ldots, dT_m.$$

The relation $\mathcal{N} = L \times \nu$ implies that, together with the point θ_0, the set $\mathcal{N}$ contains the points $(\theta_{10}, \ldots, \theta_{m0}, \theta_{m+1}, \ldots, \theta_s)$ for all values of $\theta_{m+1}, \ldots, \theta_s$. Hence from the condition $E_\theta \chi = 0$; $\theta \in \Omega \bigcap \Pi$ we deduce that

$$\int \psi(T) \exp(\theta_{m+1} T_{m+1} + \ldots + \theta_s T_s) dT_{m+1} \ldots dT_s = 0$$

for all $(\theta_{m+1}, \ldots, \theta_s) \in \omega$. By the uniqueness theorem for the Laplace transform, $\psi(T) = 0$. We have further for $\theta_0 \in \Omega \bigcap \mathcal{N}$:

$$E_{\theta_0}(g\chi) = C(\theta_0) \int g(T_{m+1}, \ldots, T_s) dT_{m+1} \ldots dT_s$$
$$\times \int \chi(T) h(T) \exp(\theta_{10} T_1 + \ldots + \theta_{m0} T_m) dT_1 \ldots dT_s$$
$$= C(\theta_0) g(T_{m+1}, \ldots, T_s) \exp(\theta_{m+1,0} T_{m+1} + \ldots + \theta_{s,0} T_s) dT_{m+1} \ldots dT_s$$
$$\times \int \chi(T) h(T) \exp(\theta_{10} T_1 + \ldots + \theta_{m0} T_m) dT_1 \ldots dT_m$$
$$= C(\theta_0) \int g(T) \psi(T) \exp(\theta_{m+1,0} T_{m+1} + \ldots + \theta_{s_0} T_s) dT_{m+1} \ldots dT_s$$
$$= 0.$$

By Lemma 1, $g(T)$ is the best unbiased estimator of the function $\gamma(\theta)$.

The proof of the necessity of the conditions of Theorem 13.1 is more complicated and the following Lemmas are required. Denote by $\mathfrak{D} = \mathfrak{D}(\mathcal{T})$ the space of infinitely differentiable functions of T with compact supports lying in $\mathcal{T}$. For any function $\phi(T) \in \mathfrak{D}$ the Laplace transform

$$\tilde{\phi}(\theta) = \int \exp(\theta, T)\phi(T)dT$$

is an entire function in C^s.

Directly from Lemma 1 it follows that if $g(T)$ is the best unbiased estimator of the function $\gamma(\theta)$, $\theta \in \Omega \bigcap \Pi$ and $\chi(T) \in \mathfrak{D}$ is such that

$$\tilde{\chi}(\theta) = 0, \qquad \theta \in \Omega \bigcap \Pi$$

then

$$\tilde{g}\tilde{\chi}(\theta) = 0, \qquad \theta \in \Omega \bigcap \Pi. \qquad \ldots \quad (13.7)$$

Lemma 2. If an entire function $\phi(\theta)$ is equal to zero on $\Omega \bigcap \Pi$, then it is equal to zero on the whole set $\mathcal{N}$.

Proof. Let $\mathcal{M}$ be an irreducible algebraic variety in C^s. By a theorem of H. Whitney [2] there exists a proper subvariety $\mathcal{M}_* \subset \mathcal{M}$ with the following property : the set $(\mathcal{M} \setminus \mathcal{M}_*) \bigcap R^s$ in a vicinity of each of its points θ is a real analytic subvariety of dimension d, and the vectors grad $f(\theta)$ (where $f(\theta)$ is an arbitrary real polynomial vanishing on $\mathcal{M}$) form an $(s-d)$-dimensional space.

Now represent the variety $\mathcal{N}$ as the sum of irreducible subvarieties

$$\mathcal{N} = \mathcal{N}^1 \bigcup \ldots \bigcup \mathcal{N}^l.$$

Applying to each $\mathcal{N}^\lambda$ the theorem of Whitney, we separate in it the exclusive subvariety $\mathcal{M}^\lambda_*$. We remark that each of the sets $\mathcal{N}^\lambda \setminus \mathcal{N}^\lambda_*$ has a non-empty intersection with Ω, since otherwise $\Omega \bigcap \Pi$ would be a part of $\bigcup_{\mu \neq \lambda} \mathcal{N}^\mu \bigcup \mathcal{N}^\lambda$, and $\mathcal{N}$ would not be the least algebraic variety containing $\Omega \bigcap \Pi$. We now fix λ and choose a point $\theta \in (\mathcal{N}^\lambda \setminus \mathcal{N}^\lambda_*) \bigcap \Omega$. By the theorem of Whitney, in a vicinity of the point θ the set $\mathcal{N}^\lambda \bigcap \Omega$ is a real analytic variety of dimension d; moreover there exist real polynomials $f_1, \ldots, f_{s-d}$, vanishing on $\mathcal{N}^\lambda \bigcap \Omega$,

whose gradients are linearly independent at the point θ. The polynomials $f_1, \ldots, f_{s-d}$ vanish on $\mathcal{H}^\lambda$, since otherwise $\mathcal{H}$ would not be the least variety containing $\Omega \bigcap \Pi$. Hence there exists a complex neighbourhood U of the point θ such that $\mathcal{H}^\lambda \bigcap U$ is a complex analytic variety of dimension d. Since $\phi(\theta) = 0$ for $\theta \in \Omega \bigcap \mathcal{H}^\lambda$, it follows that $\phi(\theta) = 0$ for $\theta \in U \bigcap \mathcal{H}^\lambda$. But any complex irreducible variety is a connected analytic variety with the exception of a nowhere dense set. Hence $\phi(\theta) \equiv 0$, $\theta \in \mathcal{H}^\lambda$. Since λ is one of the numbers $1, \ldots, l$, $\phi \equiv 0$ on $\mathcal{H}$. Lemma 2 is proved.

From (13.7) and Lemma 2 it follows that for any $\chi \in \mathfrak{B}$ with

$$\tilde{\chi}(\theta) = 0, \qquad \theta \in \Omega \bigcap \Pi$$

we have

$$\tilde{g}\tilde{\chi}(\theta) = 0, \qquad \theta \in N.$$

Let $\mathcal{A}$ be the ring of all polynomials of $\theta \in C^s$ with complex coefficients. If $\mathcal{I}$ is an ideal in $\mathcal{A}$, and $\zeta \in C^s$, we shall denote by $\mathcal{I}_\zeta$ the ideal formed by the polynomials $p(\theta+\zeta)$, where $p(\theta) \in \mathcal{I}$. Consider the set $\mathcal{G} \subset \mathcal{A}$ of polynomials $p(\theta)$ for which $g(T)$ is the best unbiased estimator of the function $\gamma(\theta)$ and is also a generalized solution in $\mathcal{I}$ of the equation

$$p(D)g = 0, \quad D = \left(\frac{\partial}{\partial T_1}, \ldots, \frac{\partial}{\partial T_s}\right).$$

It is easy to see that $\mathcal{G}$ is an ideal. Denote by I the ideal in $\mathcal{A}$ formed by all polynomials vanishing on $\mathcal{H}$.

Lemma 3. If $\zeta \in \mathcal{H}$, then $I_\zeta \subset \mathcal{G}$.

Proof. Let $p(\theta) \in I_\zeta$, i.e. $p(\theta-\zeta) \in I$. Considering $g(T)$ as generalized function on $\mathcal{I}$, we have for an arbitrary function $\phi \in \mathfrak{B}$:

$$\begin{aligned}(p(D)g, \phi \exp(\zeta, T)) &= (g, p(-D)\phi \exp(\zeta, T)) \\ &= (g, \exp(\zeta, T)p(-D-\zeta)\phi) \\ &= \int g(T)\exp(\zeta, T)p(-D-\zeta)\phi(T)dT. \qquad \ldots \quad (13.8)\end{aligned}$$

Since $p(\theta-\zeta) \in I$, the function $p(-D-\zeta)\phi = p(\theta-\zeta)\tilde{\phi}(\theta)$ vanishes on $\Omega \bigcap \Pi$. But then we have

$$\int g(T)\exp(\zeta, T)p(-D-\zeta)\phi(T)dT = 0, \; \theta \in \Omega \bigcap \Pi$$

because the expression in the left-hand side is equal to $\tilde{g}\,\tilde{\chi}(\theta)$, where $\chi = p(-D-\zeta)\phi$ satisfies the condition $\tilde{\chi}(\theta) = 0, \theta \in \Omega \bigcap \Pi$. Now from (13.8) we deduce that the generalized function $p(D)g$ vanishes for the functions of the form

$$\phi \exp(\zeta, T).$$

As any function from $\mathfrak{A}$ can be represented in this form, $p(D)g = 0$ on $\mathcal{I}$, i.e. $p(\theta) \in \mathcal{G}$. Lemma 3 is thus proved.

Consider the ideal $\mathcal{I} = \{I_\zeta\}_{\zeta \in \mathcal{N}}$. By Lemma 3, each summand I_ζ belongs to $\mathcal{G}$; hence $\mathcal{I} \subset \mathcal{G}$. Moreover, each polynomial from $\mathcal{I}$ vanishes on the set $L = \bigcap_{\zeta \in \mathcal{N}} (\mathcal{N}-\zeta)$. But if $\theta \in L$, then for a certain $\zeta \in \mathcal{N}$, $\theta \in (\mathcal{N}-\zeta)$. Hence we can find a polynomial $f \in I_\zeta \subset \mathcal{I}$ such that $f(\theta) \neq 0$. Thus L is the set of common roots of polynomials from $\mathcal{I}$ and therefore is an algebraic variety.

Lemma 4. The set $L = \bigcap_{\zeta \in \mathcal{N}} (\mathcal{N}-\zeta)$ is a linear subspace, and for each point $\zeta \in L$

$$\mathcal{I}_\zeta = \mathcal{I}. \qquad \ldots (13.9)$$

Proof. Let $\zeta \in L$; if $\theta \in \mathcal{N}$, then $\theta+\zeta \in \mathcal{N}$ because $\zeta \in \mathcal{N}-\theta$. Hence

$$\mathcal{I}_\zeta = \{I_{\theta+\zeta}\}_{\theta \in \mathcal{N}} \subset \{I_\theta\}_{\theta \in \mathcal{N}} = \mathcal{I}, \qquad \ldots (13.10)$$

i.e. $\mathcal{I}_\zeta \subset \mathcal{I}$. But the set of all roots of the polynomials from $\mathcal{I}_\zeta$ is $L-\zeta$ and so from (13.10) it follows that $L-\zeta \supset L$, i.e. $L \supset L+\zeta$ for all $\zeta \in L$. Hence L is a semi-group with respect to the operation of vector addition in C^s. We shall now show that L is a linear subspace in C^s. Let Λ be the largest linear subspace contaned in L (note that L contains at least the origin of coordinates). If $\Lambda \neq L$, there exists a point $\zeta \in L \setminus \Lambda$. Now take an arbitrary point $\theta \in \Lambda$. Since L is a semigroup, it must contain the points $\theta+k\zeta$; $k = 1, 2, \ldots$.

Let $p(\theta)$ be an arbitrary polynimial vanishing on L; since $p(\theta+k\zeta) = 0$, $k = 1, 2, \ldots$, $p(\theta+k\zeta) = 0$ for all $\lambda \in C^1$. The set of all straight lines $\theta+\lambda\zeta$, $\lambda \in C^1$ forms a linear subspace Λ' spread on Λ and ζ. Since L is an algebraic variety, $\Lambda' \subset L$; i.e. Λ is not the largest linear subspace contained in L. This contradiction proves that $L = \Lambda$,

Since L is a linear subspace, it follows from $\zeta \in L$ that $-\zeta \in L$. Hence for any points $\theta \in \mathcal{N}$ and $\zeta \in L$, we have $\theta - \zeta \in \mathcal{N}$. Hence in (13.10) we have an inverted inclusion and (13.9) is proved.

Lemma 5. The set of vectors

$$\operatorname{grad} p(0), \quad p \in \mathcal{I} \qquad \ldots \quad (13.11)$$

coincides with the space $L^{\perp}$ orthogonal to L.

Proof. It is obvious that all the vectors (13.11) belong to $L^{\perp}$ and that they form a linear space. We shall show that this linear space contains all the vectors from $L^{\perp}$. Suppose this is not so; then there exists a vector $a \in L^{\perp}$ which is orthogonal to all the vectors (13.11). Let q be an arbitrary polynomial from I. For each point $\zeta \in \mathcal{N}$ the polynomial $q(\theta+\zeta)$ belongs to $\mathcal{I}$ and so the vector grad $q(\zeta)$ is orthogonal to a. Hence $q'_a(\zeta) = 0$ on $\mathcal{N}$, i.e. $q'_a \in I$. Applying this argument to the polynomial q'_a we see that $q''_a = 0$ on $\mathcal{N}$, as was to be proved. Hence at each point $\zeta \in \mathcal{N}$ all derivatives of the polynomial q in the direction of the vector a are equal to zero. This means that q vanishes on the whole line $\zeta+\lambda a$, $n \in C^1$. On the other hand, $\mathcal{N}$ is the set of the roots of all polynomials from I. Hence the set $\mathcal{N}$, together with each of its point ζ, contains the whole straight line $\{\zeta+\lambda a\}$. Hence the set $L = \bigcap_{\zeta \in \mathcal{N}} (N-\zeta)$ contains the straight line $\{\lambda a\}$, in contradiction to the orthogonality of a to L. This contradiction proves the lemma.

Lemma 6. Each polynomial $p(\theta)$ vanishing on L belongs to $\mathcal{I}$.

Proof. Select the polynomials $p_1, \ldots, p_m \in \mathcal{I}$ in such a way that the vectors grad $p_i(0)$, $i = 1, \ldots, m$ form a basis in the space $L^{\perp}$. Construct a regular system of coordinates $(w_1, \ldots, w_m)$ in a vicinity of the origin so that $w_1 = p_1, \ldots, w_m = p_m$. Since the polynomial p vanishes on the subspace $w_1 = \ldots = w_m = 0$ (i.e. on L), it can be written in the form

$$p = \sum_1^m w_i f_i = \sum_1^m p_i f_i$$

where f_i are certain functions holomorphic at the origin. By (13.9) we can also obtain the representation

$$p = \Sigma p'_i f_i, \quad p'_i \in \mathcal{I} \qquad \ldots \quad (13.12)$$

where f_i are certain functions holomorphic at an arbitrary point ζ of L. Such a representation can obviously be obtained for each point $\zeta \overline{\in} L$, since we can always find a polynomial $p' \in \mathcal{I}$ that does not vanish at ζ.

We fix now an arbitrary point $\zeta \in C^s$. Denote by R_ζ the ring of all rational functions in C^s whose denominators do not vanish at the point ζ; by $\mathcal{H}_\zeta$ we denote the ring of all functions holomorphic in ζ. As is well known, the pair $(R_\zeta, \mathcal{H}_\zeta)$ is flat (see [3]). This means that $\mathcal{H}_\zeta$ is a flat R_ζ-module [3] and that for each R_ζ-module E the natural operation $E \to E \otimes R_\zeta \mathcal{H}_\zeta$ is injective. We now take for E the factoring $R_\zeta / \mathcal{I} R_\zeta$ where $\mathcal{I} R_\zeta$ is the ideal in R_ζ formed by all the functions of the form

$$\Sigma q_j r_j \; ; \; q_j \in \mathcal{I}, \; r_j \in R_\zeta.$$

Since $\mathcal{H}_\zeta$ is a flat R_ζ-module, we have

$$E \otimes \mathcal{H}_\zeta \cong \mathcal{H}_\zeta / \mathcal{I} \mathcal{H}_\zeta$$

where $\mathcal{I}\mathcal{H}_\zeta$ is an analogous ideal in $\mathcal{H}_\zeta$. Hence we can assert that the natural mapping

$$R_\zeta / \mathcal{I} R_\zeta \to \mathcal{H}_\zeta / \mathcal{I} \mathcal{H}_\zeta$$

is injective. By (13.12) the polynomial p belongs to $\mathcal{I}\mathcal{H}_\zeta$ and hence belongs to $\mathcal{I}R_\zeta$, in view of the injectivity of this mapping. Hence,

$$q_\zeta p \in \mathcal{I} \qquad \ldots \quad (13.13)$$

where q_ζ is a certain polynomial distinct from zero in the point ζ. Consider an ideal in $\mathcal{A}$ generated by the polynomials q_ζ. By the Hilbert Nullstellensatz (see, for example, [4]) the unity of the ring $\mathcal{A}$ belongs to that ideal, i.e. $\sum_1^k h_i q_{\zeta_i} = 1$ with certain $h_i \in \mathcal{A}$. Putting $\zeta = \zeta_i$ in (13.13), multiplying by h_i and adding, we finally obtain $p \in \mathcal{I}$, which proves the lemma.

We can now complete the proof of Theorem 13.1. In the space C^s choose a rectangular system of coordinates $\theta'_1, \ldots, \theta'_s$ in such a way as to make L coincide with the subspace in which $\theta'_1 = \ldots = \theta'_m$.

In that case $\theta'_1, \ldots, \theta'_m$ vanish on L and by Lemma 6 belong to the ideal $\mathcal{I}$. Since $\mathcal{I} \subset \mathcal{G}$, in the corresponding system of coordinates in R^s we have in the domain $\mathcal{I}$ the equations

$$\frac{\partial g}{\partial T'_1} = \ldots = \frac{\partial g}{\partial T'_m} = 0.$$

These equations in the space of generalized functions mean that the function g, up to its values on a set of measure zero, is constant with respect to the variables $T'_1, \ldots, T'_m$ in the domain $\mathcal{I}$. The representation $\mathcal{N} = L \times \nu$ is obvious, because it follows from $L \subset \mathcal{N} - \zeta$ that together with the point ζ the set $\mathcal{N}$ contains the whole variety $L + \zeta$.

Theorem 13.1 is proved.

It is interesting to compare the situation with optimal unbiased estimation of parametric functions for complete and incomplete exponential families. For the former, by the Rao-Blockwell-Kolmogrov theorem, every statistic $g(T)$ depending only upon sufficient statistics is an optimal estimator of its mathematical expectation $E_\theta g$. For the incomplete exponential families, in view of Theorem 13.1, the optimality of $g(T)$ as an estimator of $E_\theta g$ means, roughly speaking, a kind of quasicompleteness with respect to some of parameters (the representation $\mathcal{N} = L \times \nu$ is analogous to the completeness). As regards the optimal estimator itself, it depends on sufficient statistics, having the same indices as the parameters with the "quasicompleteness" property.

REFERENCES

[1] A. M. Kagan, V. P. Palamodov, Conditions of optimal unbiaised estimation of parametric functions for incomplete exponential families. *Teoria Veroiatn. i primenenia*, 12, 1 (1967).

[2] H. Whitney, Elementary structure of real algebraic varieties, *Ann. of Math.*, 66, 1 (1957).

[3] J. P. Serre, Géométrie algébrique et géométrie analitique (in French), *Ann. Inst. Fourier Grenoble*, 6 (1955-1956), pp. 1–42.

[4] B. Van der Waerden, *Modern Algebra*, 1, 1947.

CHAPTER FOURTEEN

SOME PROPERTIES OF ESTIMATORS OF A LOCATION PARAMETER

The case of families of distributions depending on a parameter taking on the values in a group is of considerable interest. Even in the simplest case of a location parameter Rao [1], Rao, Kagan and the author [2] (see also [3]) obtained rather interesting results, supplemented by Kagan, which will be expounded here.

Consider a family of distributions on E_n given by densities

$$p(x_1-\theta, \ldots, x_n-\theta) \qquad \ldots \quad (14.1)$$

with respect to the Lebesgue measure, θ being a location parameter, and the class $\mathcal{U}$ of unbiased estimators A of θ, satisfying the condition

$$A(x_1+c, \ldots, x_n+c) = c+A(x_1, \ldots, x_n) \qquad \ldots \quad (14.2)$$

for all values of c. Putting here $c = -\bar{x}$ we see that any statistic $A \in \mathcal{U}$ is of the form

$$A(x_1, \ldots, x_n) = \bar{x}+A(x_1-\bar{x}, \ldots, x_n-\bar{x}). \qquad \ldots \quad (14.3)$$

This means that the class $\mathcal{U}$ consists of all estimators of the form $\bar{x}+f(x_i-x_j)$ where a statistic $f(x_i-x_j)$ depends only on the differences of observations and satisfies the condition

$$E_0 f = 0,$$

$E_\theta f$ denoting the mathematical expectation corresponding to the value $\theta = 0$.

We pass now to the theorem by C. R. Rao from [1].

Theorem 14.1. (C. R. Rao) *An estimator $A \in \mathcal{U}$ is inadmissible within the class of unbiased estimators of the location parameter θ unless*

$$E_\theta(A \mid x_2-x_1, \ldots, x_n-x_1) = \theta \qquad \ldots \quad (14.3)$$

where $E_\theta(\cdot \mid x_2-x_1, \ldots, x_n-x_1)$ *is the conditional expectation given the differences* $x_2-x_1, \ldots, x_n-x_1$.

The loss function is here (and always in this lecture) quadratic and inadmissibility is related to this loss function.

To prove Theorem 14.1 we note that

$$E_\theta\{A(x_1, \ldots, x_n) \mid x_2-x_1, \ldots, x_n-x_1\} = E_0\{A(x_1+\theta, \ldots, x_n+\theta) \mid x_2-x_1, \ldots, x_n-x_1)\} = \theta+E_0\{A(x_1, \ldots, x_n) \mid x_2-x_1, \ldots, x_n-x_1\}. \quad \ldots (14.4)$$

Set

$$\chi = E_0\{A(x_1, \ldots, x_n) \mid x_2-x_1, \ldots, x_n-x_1\}. \quad \ldots (14.5)$$

By (14.4) χ is an unbiased estimator of zero. Let us construct now the estimator

$$B(x_1, \ldots, x_n) = A-\chi.$$

We have

$$E_\theta B = E_\theta A-E_\theta\chi = \theta.$$

Calculate now the variance of A in terms of characteristics of the statistics B and χ :

$$\begin{aligned} \mathfrak{D}_\theta A &= E_\theta(A-\theta)^2 = E_\theta(B-\theta+\chi)^2 \\ &= E_\theta(B-\theta)^2+E_\theta\chi^2+2E_\theta[\chi(B-\theta)]. \end{aligned}$$

Further,

$$\begin{aligned} E_\theta[\chi(B-\theta)] &= E_\theta\{E_\theta[\chi(B-\theta) \mid x_2-x_1, \ldots, x_n-x_1]\} \\ &= E_\theta\{\chi E_\theta[B-\theta \mid x_2-x_1, \ldots, x_n-x_1]\} \\ &= 0. \end{aligned}$$

Hence, $\mathfrak{D}_\theta A > E_\theta(B-\theta)^2$ unless $\chi = 0$ with probabiltity 1. In clase of $\chi = 0$ we have the relation (14.3) which proves Theorem 14.1.

Consider now the special case when $A = \bar{x}$. We have to calculate the "correction" χ defined by (14.5). Elementary calculations (see [3], p. 135) show that

$$\chi = E_0(\bar{x} \mid x_2-x_1, \ldots, x_n-x_1) = \bar{x}-\frac{\int_{-\infty}^{+\infty} \xi p(x_1-\xi, \ldots, x_n-\xi)d\xi}{\int_{-\infty}^{+\infty} p(x_1-\xi, \ldots, x_n-\xi)d\xi}. \quad \ldots (14.7)$$

The ratio of integrals in the right-hand part of (14.7) is called the Pitman estimator of the location parameter. Consider now the case of independent identically distributed $x_1, \ldots, x_n$ when the density (14.1) takes on the form

$$p(x_1-\theta) \ldots p(x_n-\theta).$$

Then we have the following theorem by three authors [2, 3].

Theorem 14.2. *If* $n \geqslant 3$ *and* Ex_j *exists then*

$$\chi = E_0(\bar{x} \mid x_2-x_1, \ldots, x_n-x_1) = 0 \qquad \ldots \quad (14.8)$$

iff $x_1, \ldots, x_n$ *have a normal distribution. This means that in case of* $n \geqslant 3$, $\bar{x}$ *is admissible estimator of a location parameter only in case of normally distributed observations.*

We see here the exceptional role of the normal distribution; Theorem 14.2 gives a characterization of it. It is the only distribution for which $\bar{x}$ is possibly an admissible estimator of a location parameter θ on the base of a sample $x_1, \ldots, x_n$.

Admissibility of $\bar{x}$ in case of normally distributed observations does not follow from Theorem 14.2 which states only that for any other distribution $\bar{x}$ is inadmissible. However, admissibility of $\bar{x}$ in case of normally distributed observations is almost trivial : it is only unbiased estimator of θ depending upon the sufficient statistic. In fact $\bar{x}$ itself is a sufficient statistic and the family of distributions of $\bar{x}$ is complete (see Chapter 2). If there exists another unbiased estimator of θ, say T, which depends only upon the sufficient statistic of the family, then $\bar{x}-T$ will be unbiased estimator of zero and since the family is complete, $\bar{x}-T = 0$ with probability 1. For any estimator Y (which depends may be not only upon the sufficient statistic) it is possible to construct the statistic $E_\theta(Y \mid \bar{x})$ which is also an unbiased estimator, depends only upon the sufficient statistic $\bar{x}$ and then to apply the Rao-Blackwell-Kolmogorov theorem (see, for example, [3, ch. 2]).

We pass now to the proof of Theorem 14.2. Suppose that

$$E_0(\bar{x} \mid x_2-x_1, \ldots, x_n-x_1) = 0 \qquad \ldots \quad (14.9)$$

for all values of $x_2-x_1, \ldots, x_n-x_1$ and let $n \geqslant 3$. Averaging both sides of (14.9) with respect to the values $x_4-x_1, \ldots, x_n-x_1$ (x_2-x_1, x_3-x_1 being fixed) we get

$$E_0(\bar{x} \mid x_2-x_1, x_3-x_1) = 0. \qquad \ldots \quad (14.10)$$

Introduce now the characteristic function $\phi(t) = E_0 \exp itx_j$; (it is independent of j, because of distribution of $(x_1, \ldots, x_n)$. From (14.10) we have

$$E_0\{\bar{x} \exp it_1(x_1-x_2)+it_2(x_1-x_3)\} = 0 \qquad \ldots \quad (14.11)$$

for any real t_1 and t_2.

As $E_0 x_j = 0$ we have

$$\phi'(t) = iE_0(x_j \exp itx_j)$$

(see, for example, [5]); now (14.11) can be written as

$$\begin{aligned} &E_0\{(x_1+x_2+x_3) \exp [i(t_1+t_2)x_1-t_1x_2-t_2x_3]\} \\ &= -i\phi'(t_1+t_2)\phi(-t_1)\phi(-t_2)-i\phi(t_1+t_2)\phi'(-t_1)\phi(-t_2) \\ &\qquad -i\phi(t_1+t_2)\phi'(-t_2)\phi(-t_1) = 0 \end{aligned}$$

In a sufficiently small vicinity of zero : $|t_1| \leqslant \epsilon$, $|t_2| \leqslant \epsilon$ the function $\phi(t)$ does not vanish and we have

$$\frac{\phi'(t_1+t_2)}{\phi(t_1+t_2)}+\frac{\phi'(-t_1)}{\phi(-t_1)}+\frac{\phi'(-t_2)}{\phi(-t)} = 0. \qquad \ldots \quad (14.12)$$

We set $\psi(t) = \dfrac{\phi'(t)}{\phi(t)}$ in this vicinity and show that $\psi(t) = at$.

From (14.12) it follows $\psi(t_1+t_2) = -\psi(-t_1)-\psi(-t_2)$; moreover we have $\psi(0) = 0$. Setting $t_1 = -t_2$ we find that

$$\psi(t_1) = -\psi(-t_1)$$

so that $\psi(t)$ is an odd function. Hence

$$\psi(t_1+t_2) = \psi(t_1)+\psi(t_2)$$

in the vicinity of zero. Since $\psi(t)$ is continuous, it has to be linear where from it follows

$$\phi(t) = C \exp at^2$$

in the vicinity of zero. As $\phi(t)$ is a characteristic function $C = 1$, $a \leqslant 0$. Then for all values of t, $\varphi(t) = \exp at^2$ (see, for example, [5]) and x_j has a normal distribution. The assertion of Theorem 14.2 is not true, in general, in case of $n = 2$.

Theorem 14.2 was recently extended by Kagan; we shall expound here his results.

We have seen that if $n \geqslant 3$ and the distribution of x_j is other than normal then the estimator $\bar{x}$ is inadmisisible and can be improved with the aid of the statistic χ depending only on $x_2 - x_1$, $x_3 - x_1$. Investigate now whether or not it is possible to improve $\bar{x}$ with the help of only one linear statistic.

Let us take two statistics $\bar{x}$ and $\sum_1^n a_j x_j$ and define a new statistic

$$Y = \sum_1^n a_j x_j - \bar{x} \sum_1^n a_j = \sum_1^n (a_j - \bar{a}) x_j = \sum_1^n b_j x_j.$$

Here

$$\bar{a} = \frac{1}{n} \sum_1^n a_j, \qquad b_j = a_j - \bar{a}.$$

As $\sum_1^n b_j = 0$, we have

$$Y = \sum_1^n b_j (x_j - \bar{x}).$$

The distribution of Y does not depend upon θ. Consider now the statistic

$$t_n = \bar{x} - E_0(\bar{x} \mid Y).$$

It is easy to see that for all θ

$$E_\theta t_n = \theta, \quad E(t_n - \theta)^2 \leqslant E(\bar{x} - \theta)^2, \qquad \ldots \quad (14.13)$$

the equality sign holding iff

$$E_0(\bar{x} \mid Y) = 0.$$

Again denote by $\phi(t)$ the characteristic function of x_j and set $\psi(t) = \dfrac{\phi'(t)}{\phi(t)}$, $\Phi(t) = \log \phi(t)$, the functions $\psi(t)$ and $\Phi(t)$ being well defined at least in the vicinity of $t = 0$,

Lemma 1. If $E_0(\bar{x}\,|\,Y) = 0$ then in the vicinity of $t = 0$

$$\sum_1^n \psi(b_j t) = 0. \qquad \ldots \quad (14.14)$$

Indeed the relation $E_0(\bar{x}\,|\,Y) = 0$ implies that

$$0 = E_0(\bar{x} \exp itY) = \frac{1}{n} E_0\Big(\sum_1^n x_j \cdot \exp it \sum_1^n b_j x_j \Big)$$

$$= \frac{i}{n} \sum_{j=1}^n [\phi'(b_j t) \prod_{k \neq j} \phi(b_k t)]. \qquad \ldots \quad (14.15)$$

If ϵ is sufficiently small then $\phi(b_j t) \neq 0, j = 1, \ldots, n$ for $|t| \leqslant \epsilon$. Dividing both sides of (14.15) by $\prod_1^h f(b_j t)$ we obtain (14.14) for $|t| \leqslant \epsilon$.

Formulate now the following conditions

I. $$\int |x|^k \, dF < \infty, \qquad k = 1, 2, \ldots \qquad \ldots \quad (14.16)$$

II. Among the numbers $|b_1|, |b_2|, \ldots, |b_n|$ there is exactly one which is maximal. Without loss of generality we may suppose that

$$|b_1| > |b_j|, \quad j = 2, \ldots, n. \qquad \ldots \quad (14.17)$$

Theorem 14.3. *Suppose that the distribution function $F(x)$ of x_j satisfies the condition* (14.16) *and is other than normal. Then the estimator $\bar{x}$ is inadmissible within the class of unbiased estimators of θ depending only upon $\bar{x}$ and a linear statistic $\sum_1^n b_j x_j$ which satisfies the condition* (14.17).

The proof of Theorem 14.3 requires several lemmas which can be easily proved by making use of the arguments due to Marcinkiewitz [4].

Lemma 2. Let x_l denote the l-th semi-invariant of $F(x)$. If

$$\sum_1^n \psi(b_j t) = 0$$

for $|t| \leqslant \epsilon$, then

$$x_k \sum_{j=1}^n b_j^k = 0$$

for $k = 1, 2, \ldots$.

The condition (14.17) implies that there exists a k_0 such that for $k \geqslant k_0$

$$\sum_1^n b_j^k \neq 0.$$

Therefore, $x_j = 0$ for $j \geqslant k_0+1$.

Lemma 3. If the semi-invariants satisfy the condition

$$|x_k| \leqslant C, \quad k = 2, 3, \ldots,$$

the following inequalities for the central moments μ_k hold

$$|\mu_k| \leqslant C^k . k\,! . \qquad \ldots \quad (14.18)$$

Pass now to the proof of Theorem 14.3. By Lemma 1 admissibility of $\bar{x}$ implies that $E_0(\bar{x} \mid Y) = 0$. Lemmas 2 and 3 show that then the inequalities (14.18) with a certain constant C hold. Further, (14.18) guarantees the analycity of the characteristic function $\phi(t)$ in the vicinity of zero and therefore in a horizontal strip for which zero is an inner point. Since all the x_j's with $j \geqslant k_0+1$ vanish $\Phi(z) = \log \phi(z)$ is a polynomial which by virtue of the Marcinkiewitz theorem (see [4] or [5]) must be quadratic. This means that $F(x)$ is a distribution function of a certain normal distribution. The author's results concerning the linear forms [6] (see also [5]) allow to weaken the conditions of Theorem 14.3).

In a recent paper [7] Rao proved a number of interesting results of the same kind.

REFERENCES

[1] C. R. Rao. Some theorems on minimum variance estimation, *Sankhyā*, 12, 1–2 (1952).

[2] A. M. Kagan, Yu. V. Linnik, C. R. Rao. On a characterization of the normal law based on a property of the sample average, *Sankhyā*, A27, 3–4 (1965).

[3] Yu. V. Linnik. Statistical problems with nuisance parameters, Providence, *Amer. Math Soc.*, 1968.

[4] J. Marcinkiewitz. Sur une propriété de la loi de Gauss (in French), *Math. Zeitschrift*, 44, 4–5 (1938).

(5) Yu. V. Linnik. *Decomposition of probability laws* (in Russian), Leningrad Univ. Press, 1960.

[6] Yu. V. Linnik. Linear forms and statistical criteria (in Russian), *Ukrainskii. Matem. Journal* 5 (1953), pp. 270–242.

[7] C. R. Rao. Characterization of the distribution of random variables in linear structural relations, *Sankhyā*, A28, 251–260 (1966).

CHAPTER FIFTEEN

MINIMAX DETECTION OF A DISCRETE SIGNAL DISTURBED BY A NORMAL NOISE

In our study of tests of composite hypotheses in situations involving nuisance parameters we were occupied up to now with the methods of constructing all similar tests (one can also find all unbiased tests, see [1]). The question of optimality for these tests was discussed in Chapter 4 but not in details and the investigation of this problem began only recently.

In this Chapter we present a rather interesting problem with many nuisance parameters for which the question of optimality in a certain sense, though very difficult for the solution, has been already studied.

Let us consider a sample of size n formed by independent p-dimensional vectors with a normal $N(\xi, \Sigma)$ distribution

$$X_1, \ldots, X_n; \quad X_j \sim N(\xi, \Sigma).$$

Here ξ is the vector of unknown means

$$\xi = EX_j$$

and Σ is the unknown positive definite variance-covariance matrix

$$\Sigma = E(X_j - \xi)(X_j - \xi)^T$$

where X_j are represented as column-vectors

$$X_j = \begin{Vmatrix} X_{1j} \\ \ldots \\ X_{pj} \end{Vmatrix}.$$

T being the symbol for transposition.

We are interested in testing the hypothesis

$$H_0 : \xi = 0 \qquad \ldots \quad (15.1)$$

which can be interpreted as the hypothesis that the discrete signal $\xi = \begin{Vmatrix} \xi_1 \\ \dots \\ \xi_p \end{Vmatrix}$ is absent and only the noise

$$X_j \sim N(0, \Sigma).$$

is present. This noise represents "the nuisance effect", i.e. the matrix Σ is a nuisance parameter.

Consider the following alternative

$$H'_\delta : n\xi^\tau \Sigma^{-1}\xi = \delta \qquad \dots \quad (15.2)$$

where $\delta > 0$ is a given number. Another type of the alternative is

$$H'_\delta : n\xi^\tau \Sigma^{-1}\xi \geqslant \delta, \quad \delta > 0. \qquad \dots \quad (15.3)$$

The statistical meaning of these alternatives will be explained later. The "physical interpretation" of them can be the following: the quantity $n\xi^\tau \Sigma^{-1}\xi$ measures the relative energy (in a sample) of the signal with respect to that of the noise. In case of a diagonal matrix Σ,

$$\Sigma = \begin{Vmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & \sigma_p^2 \end{Vmatrix}$$

we have

$$n\xi^\tau \Sigma^{-1}\xi = n \sum_1^p \frac{\xi_j^2}{\sigma_j^2}.$$

Hence the hypothesis H_δ is that of the presence of the signal ξ with given "relative energy" and H'_δ is that of the presence of a signal ξ whose "relative energy" exceeds the given number (the latter situation seems more realistic).

The statistical meaning of the hypothesis H_δ is that it is adapted to the application of a generalization of the well known Student's tests, namely the Hotelling test.

Set

$$\bar{X} = \frac{1}{n} \sum_1^n X_i, \qquad S = \sum_1^n (X_i - \bar{X})(X_i - \bar{X})^\tau$$

The Hotelling statistic is

$$T^2 = n(n-1)\overline{X}^\tau S^{-1}\overline{X}.$$

It has the following interesting property : its distribution depends only upon the quantity δ defined by (15.2). If $\delta = 0$ (i.e. if $\xi = 0$ in view of the condition $\Sigma > 0$) the statistic T^2 is therefore similar with respect to the nuisance parameter Σ. Hence we can determine a constant T_0^2 so that the Hotelling T^2 test, consisting in rejecting H_0 when

$$T^2 = n(n-1)\overline{X}^\tau S^{-1}\overline{X} > T_0^2 \qquad \ldots \quad (15.3)$$

has a fixed level α (which does not depend upon Σ). The power function of T^2 depends only upon δ and is constant for all values of the parameters ξ, Σ specified by H_δ. This makes obvious the statistical meaning of H_δ.

The study of the hypothesis H'_δ can be reduced to that of H_δ.

The statistical properties of the T^2 test were investigated by Simeiha [2], Stein [3, 4], Giri, Kiefer and Stein [5]. In spite of simplicity of the structure of the test its statistical properties remain up to now mysterious. In 1956 Stein proved in [3] admissibility of the T^2 test for testing H_0 against the alternative

$$H'_0 : |\xi| > 0$$

but we do not know whether the T^2 test is admissible against the alternatives H_δ and H'_δ, $\delta > 0$ and n being given. In view of this situation many authors were occupied with optimality in a weaker sense, namely with minimaxity. The T^2 test of a level α will be minimax if

$$\inf_{\theta \in H_\delta} E_\theta T^2 = \sup_{\phi} \inf_{\theta \in H_\delta} E_\theta \phi \qquad \ldots \quad (15.4)$$

where ϕ runs over all tests of the same level α.

Here T^2 denotes the Hotelling test (not the statistic) and the composite parameter $\theta = (\xi, \Sigma)$ satisfies (15.2). If we replace H_δ in (15.4) by H'_δ then we obtain minimaxity of the T^2 test with respect of H'_δ.

The relation (15.4) can be simplified if we take into account that the power function

$$E_\theta T^2$$

of the T^2 test (here T^2 again denotes the test and not the statistic) is constant on H_δ:

$$\inf_{\theta \in H_\delta} E_\theta T^2 = E_\theta T^2 = \beta(\delta), \quad \theta \in H_\delta. \qquad \ldots \quad (15.5)$$

Obviously,

$$\beta(\delta) \leqslant \sup_{\phi} \inf_{\theta \in H_\delta} E_\theta \phi \qquad \ldots \quad (15.6)$$

and therefore the inverse inequality

$$\beta(\delta) \geqslant \sup_{\phi} \inf_{\theta \in H_\delta} E\,\phi$$

is equivalent to minimaxity.

From the constancy of the power function $E_\theta T^2$ on H_δ it follows that admissibility of the T^2 test implies its minimaxity.

Hence, minimaxity is a weaker property than admissibility. It was investigated in the papers [5, 6], though the results obtained there are related to special cases*. Firstly Giri, Kiefer and Stein proved minimaxity of the T^2 test for $p = 2$, $n = 3$ and the alternative H_δ and later the author, V. Pliss and O. Shalaevskiĭ proved the same result for $p = 2$, $n = 4$ and the alternatives H_δ and H'_δ. It is probable that the T^2 test is always minimax and the theorems by the author, Prohorov and Shalaevskiĭ concerning the case of large values n of the sample size, on which we shall speak later, seem to confirm the conjecture but it remains still open.

The method developed by Giri, Kiefer and Stein in [4] is very profound and interesting from the analytical point of view; we shall sketch it here.

Instead of the critical region $T^2 > T_0^2$ it is more convenient to deal with the equivalent one

$$U = \frac{T^2}{T^2+n-1} > C \qquad \ldots \quad (15.7)$$

* Note by the translator. Minimaxity of the T^2 test in case of $p = 2$ and any $n \geqslant 3$ was proved by O. Shalaevskiĭ in [9].

where the constant $C \in (0, 1)$ is chosen according to a given level $\alpha \in (0, 1)$.

We may restrict ourselves to tests depending only on the sufficient statistics $(\overline{X}, S)$. Remark further that both the problems of testing H_0 against H_δ and H'_δ are invariant with respect to the group $GL(p)$ of non-singular $(p \times p)$ matrices and certainly to any of its subgroup. This means that if $g \in GL(p)$, then g transforms the four numbers $(\overline{X}, S, \xi, \Sigma)$ to $(g\overline{X}, gSg^\tau, g\xi, g\Sigma g^\tau)$, the expression $n\xi^\tau\Sigma^{-1}\xi$ forming an absolute invariant of $GL(p)$. The wellknown theorem by Hunt-Stein (see [7], p. 336) allows us to reduce the search of minimax tests to invariant tests with respect to subgroups of $GL(p)$ possessing "asymptotically invariant" probability measures. Such a sub-group is, for example, the group G_τ of all non-singular triangular matrices (with zeros above the principal diagonal).

The maximal invariant (see for the definition [7], p. 315) of the sub-group G_τ is a p-dimensional statistic

$$R = \left\| \begin{matrix} R_1 \\ \dots \\ R_p \end{matrix} \right\|$$

which is defined on the simplex

$$R_i \geqslant 0, \quad \sum_1^p R_i \leqslant 1 \qquad \dots \quad (15.8)$$

and satisfies the relation

$$\sum_1^p R_i = U = \frac{T^2}{T^2+n-1} \qquad \dots \quad (15.9)$$

which makes evident the invariance properties of the statistic T^2.

If the hypothesis H_0 is true, the probability density of the statistic R at a point $r = (r_1, \dots, r_p)$ of the complex (15.8) is

$$f_0(r) = \frac{\pi^{-p/2}\,\Gamma\left(\frac{n}{2}\right)\left(1-\sum_1^p r_i\right)^{\frac{n-p-2}{2}}}{\Gamma\left(\frac{n-p}{2}\right)\prod_1^p r_i^{\frac{1}{2}}}$$

and does not depend upon the nuisance parameter Σ.

If the alternative H_δ is true, the distribution of the statistic R depends upon a $(p-1)$-dimensional parameter Δ,

$$\Delta = \begin{Vmatrix} \delta_1 \\ \dots \\ \delta_p \end{Vmatrix}$$

where

$$\delta_i \geqslant 0, \quad \sum_1^p \delta_i = \delta$$

and is given by the density

$$f_\Delta(r) = f_0(r) \exp\left(-\frac{\delta}{2} + \sum_{j=1}^p r_j \sum_{i \geqslant j} \frac{\delta_i}{2}\right) \prod_{i=1}^p \Phi\left(\frac{n-i+1}{2}, \frac{1}{2}, \frac{r_i \delta_i}{2}\right) \quad \dots \quad (15.10)$$

Here Φ denotes the degenerate hypergeometric function (Kummer's function) which is defined for $a > 0$, $b > 0$ and all the values of x by the series

$$\Phi(a, b, x) = \sum_{j=0}^\infty \frac{\Gamma(a+j)\Gamma(b)x^j}{\Gamma(a)\Gamma(b+j)j!}.$$

The parameter spaces for the hypotheses H_0 and H_δ are compact and the function $f_\Delta(r)$ is continuous in Δ. From the classical theory by Wald [8] it follows that all minimax tests depending only on the statistic R are Bayesian. On the other hand, since the power function of the T^2 test in constant for $\theta \in H_\delta$, the known arguments (see [7], pp. 326–329) show that in this case the converse is also true : if the T^2 test is Bayesian it is minimax.

Therefore for minimaxity of the T^2 test of the hypothesis H_0 against the alternative H_δ it is necessary and sufficient that there exixts a measure λ on the simplex

$$\Gamma : \delta \geqslant 0, \quad \sum_{i=1}^p \delta_i = \delta$$

and a constant $K = K(C, \delta)$ such that

$$\int_\Gamma \frac{f_\Delta(r_1, \dots, r_p)}{f_0(r_1, \dots, r_p)} \lambda(d\Delta) \quad \begin{cases} \geqslant K \\ < K \end{cases}$$

if

$$\sum_{i=1}^{p} r_i \begin{cases} \geqslant C \\ < C \end{cases}$$

respectively. Here C is the constant from (15.7).

Ingenious transformations (see [5]) reduce this condition to the following integral equation

$$\int_{\Gamma_1} \exp\left\{ \sum_{j=1}^{p} t_j \sum_{i>j} \frac{\beta_i}{2} \right\} \prod_{i=1}^{p} \Phi\left(\frac{n-i+1}{2}, \frac{1}{2}, \frac{\beta_i t_i}{2} \right) \lambda(\beta_1, \ldots, \beta_p) d\beta_1 \ldots d\beta_p$$

$$= \Phi\left(\frac{n}{2}, \frac{p}{2}, \frac{\gamma}{2} \right) \qquad \ldots \quad (15.12)$$

where Γ_1 is the simplex $(\beta_1, \ldots, \beta_p) : \beta_i \geqslant 0$, $\sum_1^p \beta_i = 1$ and the variables $t_i > 0$ run over the set

$$\sum_{i=1}^{p} t_i = \gamma, \qquad \gamma = C\delta.$$

This is a Fredholm equation of the first type and this type is the most difficult for the investigation (we mentioned in Chapter 1 that the Fredholm equations of the first type appear in the general theory due to N. Chentzov). If there exists a solution $\lambda(\beta_1, \ldots, \beta_p)$ which is a probability density, the T^2 test is minimax; the converse is also true.

Even existence of the solution λ as an ordinary function (or as a generalized function defined on a suitable space of smooth functions) is a very difficult problem. O. Shalaevskiĭ proved recently that in case of $p = 2$ and any integer $n \geqslant 3$ the solution $\lambda(\beta_1, \ldots, \beta_p)$ of the equation (15.12) always exists, is an analytic function in the interior of Γ_1 and belongs to the space L_1 on Γ_1. However it is unknown whether this solution is non-negative.

Observe that the sample size n enters the equation (15.12) as a parameter. If n is an integer (this is the only case when the equation (15.12) has a statistical meaning) then it is possible to show that the equation (15.12) is equivalent to an ordinary linear differential equation of the order $\frac{n-1}{2}$ when n is odd and of the order $\frac{n}{2}$ when n is even.

The solution must be normed and must satisfy a finite number of conditions which fix its first n_1 moments, n_1 being a function only of n. In case of $n = 3$ (see [5]) we have a linear equation of the first order and the solution can be written explicitly. The case of $n = 4$, investigated in the note [6], leads to the equation

$$2\beta^2(\beta-1)\lambda''+\beta(-3+6\beta+\gamma\beta-\gamma\beta^2)\lambda'+\left(3\beta+\frac{3\gamma\beta}{2}-2\gamma\beta^2\right)\lambda = 0,$$

$$\beta \in (0, 1)$$

with additional conditions. The classical theory due to Fröbenius and the theory of the Riccati equations (to which this equation can be reduced) show that the solution $\lambda(\beta)$ is unique, positive and normed (hence the T^2 test is minimax for $n = 4$).

For the more general case of large values of n a new approach was proposed in the papers [10, 11]. If one considers large values of n the hypotheses H_0 and H_δ must be sufficiently close; otherwise the power function will be very close to 1. The closeness between H_0 and H_δ means that δ must be small. For large values of n and for small δ one can construct (see [10, 11]) an approximate solution λ_n of the equation (15.12), which is normed, positive and satisfy (15.12) rather closely. This leads to the following theorem (see [11]).

Theorem 15.1. *For the H. Hotelling T^2 test we have*

$$\sup_{\phi} \inf_{\theta \in H_{\delta_n}} E_\theta \phi(X) - \inf_{\theta \in H_{\delta_n}} E_\theta T^2 = 0(\exp - n^{1/4}) \quad \ldots \quad (15.13)$$

uniformly for $\delta_n > 0$. It goes without saying that

$$\inf_{\theta \in H_{\delta_n}} E_\theta T^2 = E_\theta T^2 \qquad \text{for all} \quad \theta \in H_{\delta_n}.$$

Hence, for large values of n the T^2 test approches the minimax property rather rapidly, it if it is not exactly minimax.

A new method of investigating the question is to consider the equation (15.12) for the complex values of n. Though it has no statistical meaning, the combination of these arguments with Theorem 15.1 seems to lead to a progress in studying the problem. This new method of introducing the theory of complex varaiables into the mathematical statistics, which consists in considering the complex sample size, seems to be useful.

REFERENCES

[1] Yu. V. Linnik. Statistical problems with nuisance parameters, Providence, *Amer. Math Soc.*, 1968.

[2] J. Simeiha. An optimum property of two statistical tests, *Biometrika*, **32**, 1 (1941).

[3] C. Stein. The admissibility of Hotelling T^2-test, *Ann. Math. Stat.*, **27** (1956), pp. 616–623.

[4] C. Stein. On tests of certain hypotheses invariant under the full linear group (abstract), *Ann. Math. Stat.*, **26**, (1955), p. 760.

[5] N. Giri, J. Kiefer, C. Stein. Minimax character of Hotelling T^2-test in the simplest case, *Ann. Math. Stat.*, **34**, 4 (1963).

[6] Yu. V. Linnik, V. A. Pliss, O. V. Shalaevskii. Contribution to the theory of Hotelling test (in Russian), *Doklay Akad. Nauk SSSR*, **168**, 4 (1966).

[7] E. L. Lehmann. *Testing Statistical Hypotheses*, N.Y., Wiley, 1959.

[8] A. Wald. *Statistical Decision Functions*, N.Y., Wiley, 1950.

[9] O. V. Shalaevskiĭ. Minimax property of Hotelling T^2-test (in Russian), *Zapiski nauchnyh seminarov Leningradskogo otdelenia Mathem. Instituta*, vol. **13** (1969), pp. 138–182.

[10] Yu. V. Linnik. Approximately minimax detection of a vector-valued signal in presence of the Gaussian noise (in Russian), *Teoria verojatn. i primen.*, **11**, 4 (1966).

[11] Yu. V. Linnik, Yu. V. Prohorov, O. V. Shalaevskiĭ. Contribution to the theory of approximately minimax detection of signals in presence of the Gaussian noise (in Russian), *Teoria verojatn. i primen.*, **12**, 3 (1967).

AUTHOR INDEX

Basu, D. 3, 5, 7
Blackwell, D. 94, 97
Boltzmann, L. 1
Breny, H. 55, 60

Cartan, H. 9, 12, 13, 32, 34
Chentzov, N. 1, 5, 109
Craig, A. 14
Cramer, H. 8

Dantzig, G. 26, 28
Doetsch, G. 40

Feigelson, T. 9, 14
Fisher, R. A. 3

Giri, N. 105, 106, 111

Hoeffding, W. 68, 76
Hogg, R. 14
Hörmander, L. 32, 34, 35, 40, 85

Kagan, A. 4, 5, 66, 85, 87, 94, 95, 99, 101
Kiefer, J. 105, 106, 111
Kolmogorov, A. N. 94, 97

Lehmann, E. L. 2, 5, 7, 14, 28, 34, 43, 67, 70, 76, 83, 111
Linnik, Yu. V. 5, 19, 28, 34, 40, 43, 52, 60, 66, 76, 95, 101, 102, 111
Lucacs, E. 5

Marcinkeiwitz, J. 100, 101

Palamodov, V. 25, 26, 28, 29, 32, 33, 34, 35, 36, 40, 41, 42, 46, 69, 70, 71, 76, 77, 80, 83, 85, 87, 94
Pliss, V. 106, 111
Poincare, H. 1
Prohorov, Yu. V. 106, 111

Rao, C. R. 94, 95, 97, 101, 102
Romanovski, I. 65, 66

Scheffé, H. 7, 14
Serre, J. P. 94
Shalaevskii, O. 4, 5, 66, 106, 109, 111
Shohat, J. A. 14
Simeiha, J. 105, 111
Stein, C. 26, 28, 80, 83, 105, 106, 107, 111
Sudakov, V. 65, 66
Sverdrup, E. 27, 28

Tamarkin, J. D. 14
Van Der Wearden, 94

Wald, A. 24, 53, 55, 60, 108, 111
Whitney, H. 89, 94

Zinger, A. 5

SUBJECT INDEX

Analytical Sheaves, 9

Bartlett-Scheffé test, 43
Behrens-Fisher problem
  problem, 22
  test, 28
  similar randomised test, 32
  non-randomised test, 41
  Bartlett-Scheffé test, 43
  homogeneous similar tests, 45, 49, 53
  Fisher-Welch-Wald test, 55
  non-randomised homogeneous similar test, 61

Characterization of normality, 97
Chentzov Morphism, 1
Cotest, 30
Cotest ideal, 31

Distinguishability of distributions, 67

Exponential family
  complete, 1
  incomplete, 2

Fisher-Welch-Wald test, 55
Functions
  verifiable, 24, 66
  sequentially verifiable, 24, 66, 80
  C-verifiable, 70, 71
  R-verifiable, 80

Holomorphic connectedness, 11, 13
Hotelling's T^2
  admissibility, 105
  minimaxity, 105
  invariance, 105, 110
  Bayesian property, 108
Hunt-Stein theorem, 106
Hyperwave operator, 43
Hypotheses
  linear, 22
  linear nonhomogeneous, 22
  linear generalised, 23
  composite, 23

Kummer's function, 108

Location parameter, 95

Minimax detection, 103
Moments and ϕ-moments problem, 10, 11, 12

Nuisance effect, 104
Nuisance parameter, 21

Palamodov's theorem, 29
Parameters
  natural, 2
  complexification, 3
  linking, 4
  nuisance, 21
Pitman estimator, 97
Polynomials
  verifiable, 25
  sequentially verifiable, 25
  invariant verifiable, 69
  C-verifiable, 77
Pre-cotest, 30

Rao's theorem, 95

Similar statistics, 7, 26
Sufficient statistics
  definition, 3
  partial, 4
  completeness, 7
  bounded completeness, 7

Tests
  optimal, 28
  ϵ-optimal, 28, 39, 40
  null regularity, 46, 50
  upsetting, 56

Transformations
  affine, 8
  homothetic, 8

Unlinking of statistics, 15

Wald's axioms, 53